MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT INTREX

SEMIANNUAL ACTIVITY REPORT

15 March 1968 to 15 September 1968

PR-6

15 September 1968

CAMBRIDGE MASSACHUSETTS

ACKNOWLEDGMENTS

The research reported in this document was made possible through the support extended the Massachusetts Institute of Technology, Project Intrex, under a grant from the Carnegie Corporation, under Research Grant NSF GN-774 from the National Science Foundation and the Advanced Research Projects Agency of the Department of Defense, and under a grant from the Council on Library Resources, Inc.

TABLE OF CONTENTS

I.	INTRODUCTION	page	1
II.	RESEARCH AND DEVELOPMENT ACTIVITIES (Electronic Systems Laboratory)		3
	A. STATUS OF THE PROGRAM		3
	B. THE COMPUTER-STORED AUGMENTED-CATALOG PROGRAM		5
	1. Augmented-Catalog Inputting		5
	2. Storage and Retrieval		13
	3. The Display-Console System		37
	C. THE TEXT-ACCESS PROGRAM		44
	1. Summary		44
	2. System Design		45
	3. Storage and Retrieval Device		46
	4. Transmission Subsystem		48
	5. Transient-Display Terminal		50
	6. Film Terminal		51
	7. Text-Access System Control		55
	8. Text-Access Software		60
III.	PROJECT INTREX STAFF		67
IV.	CURRENT PUBLICATIONS		68
V.	PAST PUBLICATIONS		69

PROJECT INTREX

Activity Report

I. INTRODUCTION

Three years have passed since the plans for Project Intrex were formulated at Woods Hole. We have worked in the directions that were set by the Planning Conference, and we have made good progress. As in most research undertakings, the pace has been less heroic than the vision of the planners. We had hoped to achieve the major goals of the plan by 1968, but it now seems unlikely that we shall reach that stage before 1970.

The first factor in modifying the dynamics of the enterprise was the decision by M.I.T. to pursue Project Intrex within the normal academic environment, rather than establish a special activity outside the regular university structure. We realized that this course would make for a slower start, but we felt that the growing involvement of faculty and students along with the research staff would more than compensate for the initial delay. Our observations of the development of the Intrex team in the Electronic Systems Laboratory have reinforced this view.

A second factor in limiting the scale of our operation has been the difficulty of attracting financial support. The funding of Project Intrex has not materialized with the speed or in the magnitude of our expectations. Where we have hoped for a budget of close to $3 million for the next two years, we are at the present time assured of no more than half that amount.

In spite of this lowering of our sights, we have derived some encouragement from our negotiations for funds. Both our principal sponsors appointed strong advisory committees to review the operations of Project Intrex. We thus had an opportunity of testing our own views of the project's past and future against the judgments of a number of highly qualified observers. We learned a great deal from these discussions, and we were pleased, of course, with the generally favorable reactions.

The activities that were reviewed by these committees are going forward in two places at M.I.T. In the Electronic Systems Laboratory, under the direction of Professor J. F. Reintjes, a team of librarians and engineers is engaged in a twofold attack on the access problem. For bibliographic access, a digitally-encoded, machine-manipulated, "augmented" catalog is under investigation. The task of preparing entries for 10,000 documents is about half completed. A user-oriented program for consulting this catalog is under test. A display console is in the final stages of construction.

For full-text access to the documents that have been cataloged, an experimental transmission system is being developed. Microfiche copies of the documents are scanned and transmitted over a wide-band circuit to a remote display station where the text is reproduced either on storage cathode-ray tube or on a duplicate microfilm.

The report which follows brings the story of these experiments up to date and provides a detailed technical record.

In the Engineering Library, Charles Stevens, of the Project Intrex staff, and Rebecca Taggart, Librarian, are supervising the preparation of an operational environment in which regular library users can experiment with the facilities of Project Intrex.

The physical reconstruction and expansion of this oldest unit of the M.I.T. library system is in full progress and is to be completed by next spring.

The books, journals, and other materials that comprise the collection of this library will be divided into an active collection of limited size and a storage collection of less frequently-used materials. The contents of the active collection will be constantly changing as new materials are added and older materials transferred to the storage collection. Access to the high-priority materials in the active collection will thus not be impeded by the growing number of obsolescent documents. Microfilm and paper duplication will be used extensively to supplement direct access to the documents.

These procedures in combination with the Intrex facilities will provide experience with a blend of traditional and advanced experimental operations that will be characteristic of university libraries in the next decade.

Carl F. J. Overhage
Cambridge, Massachusetts
15 September 1968

II. RESEARCH AND DEVELOPMENT ACTIVITIES (Electronic Systems Laboratory)

A. STATUS OF THE PROGRAM

Professor J. F. Reintjes

The computer-stored augmented-catalog part of our Experimental Library Storage and Retrieval System is now working satisfactorily for demonstration purposes. Catalog look-up requests can be made by specifying subject terms, authors and/or titles. In addition, the initial group of documents thus retrieved may be narrowed by requiring that specified matches be made on additional fields of the catalog records. It is also possible to obtain a printout of the information contained in any or all of the fields associated with each catalog record.

Included in the Demonstration System is a User's Guide. This Guide, which provides detailed instructions on methods of interrogating the catalog, is available online as computer-stored information or offline as a hard-copy manual. We intend to determine receptivity of users to each form. The Guide has been prepared with an awareness of the need to make it easy for new users to become acquainted quickly with operating details of the system.

For persons who are already familiar with the System, there is available a short form of the dialog for making requests and receiving responses. The short form substantially reduces the time consumed in inputting and outputting information.

The present System is designated a Demonstration System because the data base contains catalog information on only 48 documents. Within the next few weeks, however, we plan to insert at least 1,000 catalog records into the computer. We should then be in a position to satisfy the requirements of persons who have a bona fide need for the information thus stored. Further increases in the data base will be made as we gain experience with the formatting process associated with the inputting procedure.

The Demonstration System was operated successfully during July, 1968, from West Berlin during a two-week Conference held at the Technical University of Berlin on the subject The Computer and the University.

The remaining two parts of our library system, the augmented-catalog console and the full-text-access equipment are nearing completion and should be functioning early in the next reporting period. Recent emphasis has been on the software and on the design and construction of the interface and control logic associated with these items. A substantial amount of logic circuitry is required between the buffer/controller of the catalog console and the Central Time-Sharing Computer and between the buffer/controller and the text-access equipment in order to carry out the various sequencing operations automatically. The logic circuitry is now being constructed. The software package for the buffer/controller has been completed and is being debugged.

To date approximately 5,000 journal articles have been photographed and placed on microfiche. We are, therefore, prepared to make available the full text of the documents being cataloged as soon as the text-access equipment becomes operational.

B. THE COMPUTER-STORED AUGMENTED-CATALOG PROGRAM

1. AUGMENTED-CATALOG INPUTTING

Staff Members

Mr. A. R. Benenfeld
Mrs. S. K. Escudier
Mrs. E. J. Gurley
Mrs. S. F. Lage
Miss L. T. Lee
Mrs. J. R. Marshall
Miss V. A. Miethe
Miss S. P. Niessen
Professor J. F. Reintjes
Miss M. E. Sullivan

Cataloger Assistants

Mrs. A. M. Davis
Mrs. C. U. Koss

Graduate Students

Mr. N. A. Clark
Mr. M. S. Hsu
Mr. R. Ray

Undergraduate Students

Mr. S. D. Alexis
Mr. M. P. Bruce-Lockhart
Mr. S. C. Chamberlain
Mr. D. R. Cherry
Mr. L. A. Distaso
Mr. H. D. Feldman
Mr. J. F. Kaar
Mr. R. C. Lufkin
Mr. S. M. Nierman
Mr. C. T. Pynn
Mr. J. T. Rothman
Miss S. W. Teller
Mr. T. F. Wagner

SUMMARY

Literature from three additional research areas in materials science and engineering is being added to the data base. Changes have been made in the specifications for two data elements; the coding notation for special characters has been expanded. Modifications to the indexing process have reduced the extent of word redundancy in a set of index terms. A study of the current economic feasibility of online teletypewriter-terminal keying with respect to offline paper-tape keying of cataloging data shows the online mode to be twice as expensive as the paper-tape mode; the online mode also is sensitive to the particular computer system environment. A thesis study of the effects of indexer experience over time on the subject-indexing time per document page presents evidence of an average learning period of three months' duration for librarians and two months' duration for students. Average processing times for current operations are tabulated. As of September 1, 1968, 6428 documents have been indexed.

DATA BASE AND LITERATURE SELECTION

Previously the literature from only two research areas, which reflected the interests of two research groups in materials science and engineering, had been selected for the augmented-catalog data base. Literature from additional research areas has now been added but on the basis of modified criteria. The interests of the two initial research groups are being considered as dual cores. Additional research groups must overlap in their interests with one or the other of the core groups, in addition

to meeting other criteria for the selection of a group (see Semiannual Activity Report for 15 March 1967). In this manner the number of potential system users is increased and, at the same time, the adequacy of the data base is increased without unduly enlarging that base. Three additional research groups have been selected to participate in the catalog experiments. The research groups in the two core areas are:

Core A.	1.	Radiofrequency, microwave, and optical spectroscopy of liquids and solids
	2.	Microwave and quantum magnetics
Core B.	1.	High-temperature metallurgy
	2.	Casting and solidification
	3.	Structural materials

We are continuing to route journal-issue tables of contents to two members of each research group; in return for this current-awareness service, they indicate articles which they consider important to their area. All such articles, plus any obvious oversights caught by a librarian, are accepted for the data base. In addition to the journal literature, the data base includes conference proceedings in the core-subject areas, and the technical reports and theses issued by the participating research groups. When a sufficient data base of this prime literature has been established, other literature forms will be added to round out the catalog.

PROCESSING

No significant changes have been made in the workflow; it remains essentially as described in the preceding Semiannual Activity Report. However, the addition of the literature of interest to three new research groups led to several minor changes in the procedures associated with distribution of tables of contents and with microfilm initiation.

About 4500 documents that had been cataloged before the tolerances in the current microfilming operation were established have now been filmed for microfiche preparation. The access numbers for the documents stored on these fiche will be added later to the machine-stored catalog records through an update operation.

Some documents selected for the data base will not be filmed for microfiche text storage. Such documents include those whose length exceeds 60 pages (or one fiche, there being 60 frames per fiche), those of a temporary nature (such as preprints), and those of very poor graphic quality. However, users will have guaranteed hard-copy text access to these documents.

As of September 1, 1968,

6428 documents have been indexed,

6320 records have been reviewed,

5295 records have been keyed,

485 files have entered the computer and codes converted from Flexowriter to ASCII codes,

454 files have passed through the first proofreading and editing process,

212 files have passed through three or more proofreading and editing processes,

400 files have been certified for further processing into the stored data base.

INDEXING

The student-indexer program, described in detail in the preceding Semiannual Activity Report, has continued to be quite satisfactory. Ten students were employed on a part-time basis during the Spring semester. During the summer, five students were employed; two of these students worked on a full-week basis.

The costs associated with computer storage of data are high. Temporary storage for processing and final-storage requirements are particularly large for the subject-index terms. These terms are stored in the main catalog file as full terms and in the subject/title inverted file both as stemmed-word full terms and as phrase-decomposed and stemmed individual words. In an effort to reduce the processing and storage costs without sacrificing the basis of our experiments, we have more closely examined the indexing process in light of our software-retrieval capabilities. As a result, several modifications have been made to the indexing. The modifications eliminate excessive redundancy within a set of index terms; at the same time, the basic design of the indexing and retrieval processes remains unchanged. It may be helpful to review the nature of the indexing before discussing the specific modifications to it.

The set of index terms describing a document are freely chosen combinations of phrases based upon the text of a document. Each term is structured to provide sufficient contextual expression such that the term stands by itself. This is often achieved by intensifying text phrases. Each index term is weighted. As explained in preceding reports, there are two kinds of weights. One type of weight reflects the relative proportion of a document devoted to a discussion of the represented concept. Here, a weight of 1 signifies that the term is descriptive of most of the content of an entire document; a weight of 2 indicates that the term characterizes a section of the document; and a weight of 3 refers to a term that describes a small fraction of the document. The other kind of weight is a special-purpose weight assigned irrespective of extent of discussion of the topic designated by the term. The special weight of 4 is assigned to terms

representing mathematical tools, instrumental tools, materials, or applications; the special weight of 0 (zero) is reserved for terms that are generic to the subject matter of a document. Previously, an index term might have been eligible to receive two weights, a weight from the regular series indicating the extent of discussion, and if applicable, a special series weight. Whenever dual weights were assigned, the entire index term was repeated.

At present, the software search and relevancy algorithms do not utilize the assigned and stored weights, and so it is redundant to assign more than one weight to a term. In the future, the software will utilize these term weights in the determination of the relevancy of an index term associated with a particular document to a retrieval-request term. In this case, assignment of more than one weight to a term is still redundant because only one of the weights, the higher ordered one, will be considered in determining the priority of the term. The practice of assigning two weights to the same term has been dropped from the indexing process. Because the regular weights have been given greater emphasis than the special weights, the definition of the special-purpose weights has been revised such that they are now applied only to terms which are not eligible for a regular weight. In essence, this means that special-purpose weights are now assigned only to terms that are not discussed, per se, in a document.

In addition, the phrase-decomposition retrieval programs also make redundant any index term which appears as a string within any other index term that has the same or a higher ordered weight. Consequently, for example, the creation of an index term consisting of the name of a material has been retained only if that name does not also appear within another term with a higher ordered weight.

These two modifications, the elimination of dual-weighted terms and the elimination of terms also appearing as a string within another term, do cause some loss of information. A user will not be able to narrow his subject-search requests solely on the basis of a single weight without losing pertinent references. However, other searching strategies can overcome this problem because the words themselves have been retained in the index; only word redundancies have been eliminated.

In another modification, separate terms which have similar wording, which are logically equivalent, and which have the same assigned weight, are now combined into a single index term. This single term has compound-subject or object sentence elements. This procedure eliminates repetition of words common to the separate index terms. For example, the separate terms "optical-dispersion curves for barium titanate (3)" and "optical-dispersion curves for strontium titanate (3)" would now be combined into the single term "optical-dispersion curves for barium titanate and strontium titanate (3)."

With the elimination of these redundancies, the indexing and the software which operates on it are in better harmony; in addition, computer-storage space and cost are conserved.

DATA ELEMENTS AND CODING

Additional codes have been established for Field 5, Access Number, which indicate the availability and location of guaranteed hard-copy text access to those documents not stored on microfiche.

The specifications for Field 66, Level of Approach, have been rewritten. This field suggests the professional level of the author's intended audience, or the minimum general level of understanding required to comprehend the subject of a document. The number of levels has been reduced from seven to a more easily identifiable set of three, namely, professional level, undergraduate level, and below college level.

The coding notation for simultaneous machine-readability and human recognition of characters not on our Flexowriter keyboard, called special characters (see the Semi-annual Activity Report of 15 September 1967), has been expanded. The string of characters in a superscript or subscript notation is, itself, considered a special character. Previously, we could not encode such a string if it contained a special character in a nonleading position. Square brackets now demarcate nonleading special characters. For example,

$$\chi = A(T - T_c)^{-\gamma}$$

becomes

chi = A(T - T*sub c*)*sup -[*gamma*]*

This expanded notation also permits us to code nested special characters as, for example, a superscript to a superscript.

EVALUATIONS

The average processing times for current workflow operations are tabulated in Fig. B-1. Other characteristics of these processes and of the catalog records are also listed in Fig. B-1. It takes an average of 68.1 man-minutes to catalog, review, key, and first-proofread a catalog record. In current processing, 80 percent of the files cycle through two edit loops (once for editing and once for certification); 20 percent of the files have errors to be corrected in a second online editing step and so these files are cycled through three edit loops. (A file consists of 10 catalog records.) About half of the 11.2 average number of errors per file detected during a first proofreading are typographical errors; cataloging errors comprise the remainder. As the number of errors per file decreases, the ratio of total computer-time consumed to console-time consumed in online editing increases. The ratio is 1/36 for the first online edit of a file of 10 records and it is 1/6 for online certification of an error-free file.

We can gain further perspective in this analysis by considering the size of a logical catalog record keyed for computer input. There are an average 2461 keystrokes per record, exclusive of carriage returns, tabs, and case shifts, or 410 six-character English words per record. There are 16.4 index terms per record, averaging 7.7

CURRENT AVERAGE PROCESSING TIMES
and
ASSOCIATED AVERAGE CHARACTERISTICS

Process		Average Time
		(man-minutes/record)
Indexing		28.0
Descriptive Cataloging (by Librarian)		4.7
Review		10.3
Keying plus Descriptive Cataloging (by Typist)		16.9
		(man-minutes/file)
Proofreading (by Librarian)		82.0
On-Line Editing		20.7
Second Proofreading (by Typist)		11.5
Second On-Line Editing	20% of the files	4.1
Third Proofreading (by Typist)		1.7
On-Line Certification		1.0

Associated Characteristics	Average
Computer Time/Console Time Ratio	
First On-Line Edit	0.028 (1/36)
Second On-Line Edit	0.074 (1/14)
On-Line Certification	0.17 (1/6)
Errors Caught During	(per file)
First Proofreading	11.2
First Editing	0.33
Second Proofreading	2.0
	(per record)
Keystrokes	2461
English Words	410
Index Terms	16.4
English Words/Index Term	7.7
	(pages)
Document Length	5.3

Fig. B-1 Average Processing Times of Current Workflow Operations and Associated Average Characteristics (January - April, 1968). One file contains ten records.

English words per index term in length. At present, the data base consists almost entirely of journal articles or conference papers, and in the survey sample these documents averaged 5.3 pages in length. The set of index terms comprises about 1/3 the total size of a catalog record, an abstract or excerpt comprises about 2/5 of the record size, and all other applicable data elements account for the remainder of the catalog record.

We have made a study of the economic feasibility of online teletypewriter-terminal keying with respect to our normal offline paper-tape keying of the cataloging data. Six files of catalog records were keyed online by a typist working at an IBM 2741 terminal connected to the 7094 computer. The online input was cost-compared to that for another six files input by the paper-tape medium. The analysis of each input mode included all of the computer operations through the generation of a first printout. The data are summarized in Fig. B-2; they are valid only for the current (March, 1968) M.I.T. Compatible Time-Sharing System (CTSS) environment. The unit-cost figures for typist salary (including overhead), and keying machine rental costs, are pro-rated over 20 working days per month, seven hours per day. Materials associated with online input are negligible in cost and there is no direct charge to us for printout. The analysis shows that the cost of online input of data to the augmented catalog in the current CTSS environment is $41.07 per file of ten records whereas the cost of paper-tape input is $23.46 per file. The cost differential is due almost entirely to the computer time and cost required to process data from the online terminal (205.4 seconds versus 6.4 seconds). All other operations in the two input modes are nearly equivalent in time and cost. Two computer programs (designated EDA and QED) were used in this experiment. Each program operated on a part of the file set which was input online. There was essentially no difference in the total computer time expended with each program. The ratio of computer time to console time consumed was 0.018 for EDA and 0.016 for QED. However, the components of the total computer time are program sensitive. There is a trade-off between processing time (the time for programs to operate on data in core) and swap time (the time to move programs and data into and out of core in a multiprocessing system). The ratio of processing time to swap time is 0.43 for EDA and 0.27 for QED.

As the hardware associated with the Intrex System becomes available, and as the computing-system environment changes, online input will be restudied; particular attention will be given to the effects of using a local buffer/controller which is, in turn, connected to the larger computer system.

In another study, Mr. Richard Lufkin, in fulfillment of his Bachelor of Science thesis requirement at M.I.T., made determinations of the effects of indexer experience over time on the subject-indexing time per document page. The data are scattered but seem to indicate an average learning period of three months' duration for librarians and of two months' duration for students. At the end of this learning period, indexing time per document (averaged over all indexers) levels off at six to eight minutes per

COST ANALYSIS OF ON LINE INPUT AND
PAPER TAPE INPUT

Operation (Unit Cost)	On-Line Input		Paper Tape Input	
	Time per File	Cost per File	Time per File	Cost per File
Keying Operator	(minutes)		(minutes)	
Preparation	27.0		30.0	
Log In	2.4			
Input	174.3		164.2	
Total Keying ($0.095/min.)	203.7	$19.35	194.2	$18.45
Computer	(seconds)		(seconds)	
Log In Subtotal	8.4			
Processing	7.1			
Swap	1.3			
Input Subtotal	205.4		6.4	
Processing	54.1		3.4	
Swap	151.3		3.0	
Magnetic Tape Generation, Printout and Miscellaneous Operations Subtotal	6.5		8.0	
Processing	5.6		6.8	
Swap	0.9		1.2	
Total Computer ($0.083/sec.)	220.3	$18.28	14.4	$1.20
Computer Operator ($0.055/min.)	(minutes) 3.2	$0.18	(minutes) 7.0	$0.39
Keying Machine Rental	(minutes)		(minutes)	
IBM 2741 and associated equipment ($0.016/min.)	203.7	$3.26		
Flexowriter 2303 ($0.014/min.)			194.2	$2.72
Materials				$0.70
TOTAL COST		$41.07		$23.46

Fig. B-2 Cost Analysis of On-Line Input and Paper Tape Input of Catalog Data to a 7094 Computer Operating in the M.I.T. CTSS Environment (March, 1968). One file contains ten records.

page from an initial high of ten to twelve minutes per page. In another phase of this thesis, the average indexing time per page was found to be influenced by document-related parameters. There were observable differentials of the per-page indexing time among documents grouped by format. For example, the ease of indexing articles decreases in this order: conference paper, regular journal article, letter-type journal article.

2. STORAGE AND RETRIEVAL

Staff Members

Mr. C. E. Hurlburt
Mr. P. Kugel
Mr. R. L. Kusik
Mr. R. S. Marcus
Mr. M. K. Molnar
Professor J. F. Reintjes
Professor A. K. Susskind
Mr. H. F. Vandevenne (visiting)

Graduate Students

Mr. R. Goldschmidt
Mr. W. Kampe
Mr. T. Welch

Undergraduate Students

Mr. T. Lin
Mr. K. Pogran
Mr. W. Robinson
Mr. R. Voss

SUMMARY

As described in preceding Semiannual Activity Reports, the implementation of our experimental augmented-catalog storage-and-retrieval computer programs has been scheduled in a research program of three phases:

Phase I: A restricted system for use by the Intrex programmers in investigating various techniques of file organization, storage, and retrieval.

Phase II: A more complete system, based on the present M.I.T.-modified IBM-7094 time-sharing computing system (CTSS), for conducting experiments on storage and retrieval in the context of our 10,000-document augmented catalog and a selected group of users.

Phase III: An expanded version of the Phase-II system to be implemented on next-generation time-sharing computers at M.I.T.

The Phase-I activity has been completed. To date, very little Phase-III activity has been undertaken since the anticipated new time-sharing systems will not be available to Intrex as soon as expected. Therefore, Intrex experiments will be conducted for the time being on the present M.I.T. IBM-7094 time-sharing system with the Phase-II system.

The Phase-II system has undergone rapid development in this reporting period and we now have a working version which has been termed the "Demonstration System". This Demonstration System includes most of the basic retrieval functions and

user-system dialog features that have been specified in previous Semiannual reports for our initial Phase-II system, but the data base upon which it operates is small-- 48 documents. This system has not only permitted us to demonstrate Intrex -- demonstrations have been given via telephone-communication links at a recent conference in West Berlin -- but also to experiment with actual user reaction to system operation.

Our next major goal is to expand the online data base to sufficient size (perhaps 1000 documents) so that meaningful experiments can start with users who have a real need to extract information from the system. At the same time we shall be evaluating and improving system performance and preparing for the incorporation of the display-console and text-access systems.

In addition to the above work, which is intended as a basis for experiments in the near future, a study of a storage system for a much larger catalog is under way.

DESCRIPTION OF DEMONSTRATION SYSTEM

The Demonstration System permits the user to search the data base for references to documents by specifying subjects, authors and/or titles. The user may then make a selection among the documents thus retrieved by requesting that they contain specified information in certain fields of the documents' catalog records. Finally, he may request that the information contained in any or all of the catalog fields be printed out.

In order to illustrate the nature of the Demonstration System more concretely, a sample user-system dialog is given in Fig. B-3. This dialog is reproduced from the typewritten copy prepared on an IBM 2741 teletypewriter console attached to the time-sharing system (CTSS). For illustrative purpose of this report each user statement is flagged by a number and the letter U in the left margin. Similarly, system messages are flagged by numbers and the letter S .

In his first two statements the user has logged-in to the time-sharing system (CTSS) which hosts Intrex as well as many other computer programs. Note that the user's second statement, his password to CTSS, is not printed because the 2741 is changed to nonprinting mode by CTSS for that statement to protect the security of the password. This log-in procedure will be unnecessary for individual users for consoles dedicated to Intrex use.

The third user statement "resumes" the "Intrex" system (at this point the user could have called for any other program currently resident in CTSS) and in the fourth statement the user "signs in" to Intrex by typing his name and address. The "sign-in" statement, in conjunction with the monitoring procedures (see below), provides us, as system analysts, with a means for keeping track of system use. It also serves to introduce the user to certain system procedures. For example, the user is apprised of the fact that his statement must be terminated with a carriage return (note that a more natural statement-terminator button or switch will be possible with the display console -- see Section B.3). While at present the sign-in statement merely serves to aid in monitoring system use, we anticipate future system developments whereby some

Fig. B-3 Sample Demonstration System Dialog

```
1S  TST7X5: USERS= 16, MAX= 47.
    READY.

1U  login m5806 marcus
2S  W 1315.7
    Password
2U
3S   STANDBY LINE HAS BEEN ASSIGNED
     M5806    160 LOGGED IN  08/19/68 1315.9 FROM 800277
     LAST LOGOUT WAS  08/15/68  944.4 FROM 800277
     HOME FILE DIRECTORY IS M5806 CMFL01

    DUE TO HARDWARE DIFFICULTIES, CTSS OPERATION MAY BE
    IRREGULAR.

     CTSS BEING USED IS  TST7X5
    R 6.166+1.016

3U  resume intrex
4S  W 1316.8
    Greetings! This is Intrex 1a.  Please sign in by typing your name and
    address as in the following example:

    smith, r j/mit 13-5251

    Note that your sign-in statement should end with a carriage return.
    READY
4U  marcus, r s/mit 35-406
5S  If you already know how to use Intrex, you may go ahead and type
    in commands. (Remember, each command ends in a carriage return.)
    Otherwise, for information on how to make simple queries of the
    catalog, type

    info 2

    or, to see the Table of Contents (Part 1) of Intrex-1 Guide which will
    direct you to other parts of the Guide explaining how to make more detailed
    queries, type

    info 1
    READY
5U  info 2
6S  Part 2 of Intrex 1A Guide:  Simple Queries


         To find documents in the system specify your query by subject, author,
    and/or title terms, as shown in the 3 examples below:

    subject ferroelectric transitions

    author Hess, G.B./subject helium

    title sulfurization/author Swisher

         In order to specify additional restrictions (e.g., where author comes
    from, journal, word variations not to use, etc.), see Part 9 of the Guide
    (or type info 9).  For other than standard output (document numbers, titles,
    and authors) see Part 8.  For general Intrex command format and abbreviations
    see Parts 6.1 and 6.2.
    To see Table of Contents for Intrex 1a Guide and how to use
    the Guide on line, type
```

Fig. B-3 (cont'd)

info 1

Otherwise, you may now make simple queries or use any other command.
READY

6U subject solid phase transitions

7S A search on your query SUBJECT solid phas-e transit-ions found 7 documents.
To output the catalog fields DOCUMENT NUMBER, TITLE, AUTHOR on those
documents type

go

This output will take about 15 seconds per document. You may terminate
this output at any time by hitting the ATTN key ONCE. Otherwise, you may
change your output request. For information see Part 8 of Guide or type

info 8

or change your field restriction (see Part 9.5) or make another request of
Intrex (see Part 1)
READY

7U go

8S 1. DOCUMENT NUMBER 2851

(21) AUTHOR
Hoshino, Sadao;
Shimaoka, Kohji (JA);
Niimura, Nobuo (JA)

(24) TITLE
Ferroelectricity in solid hydrogen halides

2. DOCUMENT NUMBER 3430

(21) AUTHOR
Sihvonen, Y.T.

(24) TITLE
Photoluminescence, photocurrent, and phase-transition correlations

3. DOCUMENT NUMBER 3174

(21) AUTHOR
Stanley, H. Eugene (TA)

(23) CORPORATE
Harvard University, >Cambridge<, >Mass.< (RT)

(24) TITLE
High-temperature expansions for the classical Heisenberg model,
susceptibility

4. DOCUMENT NUMBER 3124

(21) AUTHOR
Schmidt, Hartland H.;
Opdycke, Jack (TA) (JA);
Gau, Charles F. (JA)

(23) CORPORATE
University of >California<, >Riverside< (RT)

(24) TITLE
Heat capacity in the critical region of xenon

5. DOCUMENT NUMBER 3048

Fig. B-3 (cont'd)

(21) AUTHOR
Novikova, S.I.;
Shelimova, L.E. (JA)

(24) TITLE
Low-temperature phase transition in tin telluride

6. DOCUMENT NUMBER 1715

(21) AUTHOR
Nettleton, R.E.

(24) TITLE
Renormalization and the first-order ferroelectric transition in

7. DOCUMENT NUMBER 1690

(21) AUTHOR
Willens, R.H.;
Buehler, E. (JA);
Matthias, B.T. (JA)

(24) TITLE
Superconductivity of the transition-metal carbides

Output completed. Total of 7 documents found. You may now see additional output on these documents by making a new 'output' request (for information on how to do this, see Part 8 of the guide or type info 8). You may also select a portion of these documents by making a new 'infield' request (see Part 9.5). Otherwise, you may make a new search (see Part 2) or make other requests (see Part 1).

READY
8U output affiliation matchsub relevance/go
9S 1. DOCUMENT NUMBER 2851; RELEVANCE 3/3

(22) AFFILIATION
University of >Tokyo<. Institute for Solid State Physics;
University of >Tokyo<. Institute for Solid State Physics;
University of >Tokyo<. Institute for Solid State Physics

(74) MATCHSUB

phase transition at low temperature in solid hydrogen halides (0)

2. DOCUMENT NUMBER 3430; RELEVANCE 2/3

(22) AFFILIATION
>Texas< Instruments, >Dallas<

(74) MATCHSUB

(TITLE)

3. DOCUMENT NUMBER 3174; RELEVANCE 2/3

(22) AFFILIATION
Lyman Laboratory of Physics Harvard University, >Cambridge<, >
Laboratory M.I.T., >Lexington<, >Mass.<

(74) MATCHSUB

relation of high-temperature expansions for some two-dimensional
nearest-neighbor ferromagnetic interactions to phase
(T*sub c* *=/* 0) (3);

4. DOCUMENT NUMBER 3124; RELEVANCE 2/3

Fig. B-3 (cont'd)

```
(22)  AFFILIATION
          University of >California<, >Riverside<.  Dept. of Chemistry;
          University of >California<, >Riverside<.  Dept. of Chemistry/Univ
          >San Diego<.  Dept. of Chemistry;
          :National Science Foundation Undergraduate Research Participant:U
          >California<, >Riverside<.  Dept. of Chemistry

(74)  MATCHSUB

          phase-transition temperature of xenon (3);
5. DOCUMENT NUMBER 3042;  RELEVANCE 2/3, 2/3

(22)  AFFILIATION
          Baikov (A.A.) Institute of Metallurgy Academy of Science of the
          Baikov (A.A.) Institute of Metallurgy Academy of Science of the

(74)  MATCHSUB

          low-temperature phase transition in tin telluride (1);
          (TITLE)
6. DOCUMENT NUMBER 1715;  RELEVANCE 2/3, 2/3

(22)  AFFILIATION
          Sandia Laboratory, >Albuquerque<, >N.M.<

(74)  MATCHSUB

          second-order phase transition in perovskites (3);
          first-order phase transition (0);
7. DOCUMENT NUMBER 1690;  RELEVANCE 2/3, 2/3

(22)  AFFILIATION
          Bell Telephone Laboratories, >Murray Hill<, >N.J.<;
          Bell Telephone Laboratories, >Murray Hill<, >N.J.<;
          Bell Telephone Laboratories, >Murray Hill<, >N.J.<

(74)  MATCHSUB

          superconducting transition temperature as a function of compositi
                    solution (3);
          superconducting transition temperature as a function of compositi
                    solution (3);
     Output completed. Total of  7 documents found. You may now see
additional output on these documents by making a new 'output' request (for
information on how to do this, see Part 8 of the guide or type info 8).
You may also select a portion of these documents by making a new 'infield'
request (see Part 9.5).     Otherwise, you may make a new search (see Part
2) or make other requests (see Part 1).
READY
9U   infield affiliation harvard/o 71/go
10S  1. DOCUMENT NUMBER 3174

(71)  ABSTRACT
          The high-temperature series expansion of the zero-field magnetic
          *chi*/*chi**sub Curie* = 1 + *SIGMA**sub l = 1*sup *infinity**a*s
          is related to the diagrammatic representation of the correspondin
          expansion of the zero-field static spin correlation function
          **S*sub f*.S*sub g****sub *beta** presented elsewhere.  The first
general lattice constants" p*sub lx*.  The general lattice expressions are
          used to evaluate these a*sub l* numerically for three two-IINT. 1
Does the crIREADY
10U  quit
11S  Thank you for using intrex.
     R 109.583+20.800
```

history is kept to past users so that, when someone signs in, the system can take account of his previous experience to help direct him in current system use.

Intrex response to the user's sign-in statement, message 5S, is illustrative of several features of the prompting and instructional techniques employed by the system. In the first place, the user is told of the various alternative actions that he may take at any given time. Secondly, the specific form of the statement he should type to invoke one of these actions is explicitly stated, where possible. Thirdly, where it is not possible to explain the alternatives completely, the user is referred to a Guide to system use for further details.

The Guide is available both in hard-copy form and online. Selected pages from the hard-copy version are shown in Fig. B-4. The user may request that a section of the Guide be printed online by using the "info" command (see user statement 5U and system response 6S). It may be seen that the Guide itself also attempts to use the techniques of presentation of alternatives, example, and reference to more detailed information. The sections of the Guide are sized for convenient printing and viewing online. Details of the generation and maintenance of the Guide are given below in the discussion on system implementation.

The user's sixth statement initiates a search in the inverted files for documents on a given subject. Searching may also be done on title or author terms or combinations of subject, title and/or author terms. It may be noted that the form of the user's statements is a compromise between the precise but esoteric and complicated form of many programming languages and the familiar but ambiguous (and, therefore, difficult to interpret automatically) form of natural English. Command and argument names (see Part 6.2 of Guide in Fig. B-4) are simple and mnemonic. Format is kept simple with only three basic delimiters required: spaces to separate arguments from each other and from command names; slashes to separate commands; and a carriage return to terminate the statement.

In response to the user's search request Intrex replies with a message (7S) illustrative of several other features of system dialog. In the first place, the system plays back its understanding of the user's statement. Also the system indicates, by hyphenating word endings, how it has stemmed the words in the user's search specification. This is important because Intrex matches these word stems to word stems in the inverted file (see preceding Semiannual Activity Report). As a further indication of the progress of the retrieval process, the number of matching documents found in the inverted files is printed. Since the user made no special output request in statement 6U, Intrex reports the estimated time to output the standard catalog fields. This system message, then, gives feedback which may interact with the user's original intentions and expectations and allow him to redirect his search.

The points, at which the system reports to the user, have been chosen in light of the operating characteristics of the host CTSS time-sharing system. The intention is to report back soon enough so that the user experiences quick response but not to

Fig. B-4 Sections from Intrex Guide

Part 1 of Intrex 1A Guide: Table of Contents

To have a part of the Guide printed out on line use the "info" command. For example, for information on making simple queries (i.e., to print out part 2), type

info 2

PART	CONTENTS
1	Table of Contents
2	Simple Queries
3	General Remarks - How to Get Guide (printed copy)
4	Log-in to CTSS and Call Intrex
5	Typing Errors - How to Correct
6.1	Commands, Modes (LONG, SHORT), Time Checks
6.2	Command Names and Abbreviations
7	Preliminary Output
8	Final Output
9	Generalized Queries
10	Scanning Index Terms
11	Interrupting System Messages
12.1	Text Access
12.2	Library Services
13	User Comments and Questions
14	Documents in the Collection
15	The Catalog and Its Fields
16	Sample Catalog Record
17	Exit from the System

This online guide was last revised on 7/24/68.

INTREX 1A as of 24 JUL 68 PAGE 2

Part 2 of Intrex 1A Guide: Simple Queries

To find documents in the system specify your query by subject, author, and/or title terms, as shown in the 3 examples below:

subject ferroelectric transitions

author Hess, G.B./subject helium

title sulfurization/author Swisher

In order to specify additional restrictions (e.g., where author comes from, journal, word variations not to use, etc.), see Part 9 of the Guide (or type info 9). For other than standard output (document numbers, titles, and authors) see Part 8. For general Intrex command format and abbreviations see Parts 6.1 and 6.2.

Fig. B-4 (cont'd)

Part 15.73: SUBJECTS

Terms describing the subject content of the document are entered here. A term may be a single word or a phrase. Terms are commonly phrases used by the author. Eash term is assigned a 'range' as described in Part 9.2. Normally, subject specification is made in the primary part of the query as described in Part 9.

Journal articles are indexed by an average of about 20 terms. Each term has an average of about 5 words.

Part 15.74: Matching Subject Terms (MATCHSUB)

This is not an actual field in the catalog but calling for it as output (see Part 8) will cause the subject terms that matched the user's query to be included in the output. (See Part 15.75 for criterion for matching; i.e., RELEVANCE.)

If the matching term is the title the word "(TITLE)" is output. Otherwise, this field is simply a subset of field 73 (SUBJECTS).

Since this is not an actual field, an INFIELD request would be inappropriate.

Part 15.75: RELEVANCE

This is not an actual field in the catalog but calling for it as ouptut (see Part 8) will cause the estimated relevance of each matching subject term (see Part 15.74) to be included in the output.

At present, relevance is estimated as the ratio of the number of individual English words in the user subject search request that actually match (i.e., [illegible] on word stems) for a given subject term in the catalog [illegible] the total number of words in the subject search request.

The present criterion for matching depends on the number of words in the subject search request. If there are one or two words, each must match If there are three or four words, no more than one can fail to match (i.e., at least 2/3 or 3/4). If there are more than four words in the subject search request, at least three of them must match words in the given subject term.

Fig. B-4 (cont'd)

Part 3 of Intrex 1A Guide: General Remarks

Intrex 1A is an initial demonstration version of a system for automatic retrieval of information from libraries. For this demonstration the data base contains information on 48 journal articles in selected areas of Materials Science and Engineering (including spectroscopy of liquids and solids and high temperature metallurgy). The data base includes all the information usually found in card catalogs plus much additional information like abstracts, level of approach, and author's purpose. For details on the data base see Parts 14, 15, and 16 of this Guide. You may interrogate this data base to find references to documents covering broad or specific questions of interest.

How to use Intrex 1A is explained in "Intrex 1A Guide" for which this is the introduction. Contact Project Intrex at M.I.T. Room 35-416, ext. 2354, to obtain copies of this guide or other information about Intrex. This guide is also available online at the computer as explained in Part 1.

References on Project Intrex include the Semiannual Activity Reports (Call No. Z699, A1.M41.P96) and the Report of the Planning Conference 1965 (Call No. Z699.A1.P712). They are available at the Engineering Library

INTREX 1A as of 24 JUL 68 PAGE 6

Part 6.2: List of Commands

A list of commands, special arguments (and abbreviations) and Part of Guide in which explained is given below:

NAME	ABBREVIATION	PART
info		1
subject	s	2
author	a	2
title	t	2
time		6.1
on		6.1
off		6.1
long	l	6.1
short	sh	6.1
go	g	7
seematch	sm	7
output	o	8
standard	+	8
all		8
range	r	9.2
not		9.2
major	1	9.2
secondary	2	9.2
minor	3	9.2
tool	4	9.2
generic	0	9.2
infield	in	9.4
scan		10
library	lb	12.2
comment	c	13
exit	ex	17
quit	q	17

report so often that the user is forced into unnecessary additional responses of his own, thereby forcing up not only the overall real time but also the "swap" time in the computer which has to read back the Intrex programs on each interaction. Of course, the incorporation of the buffer controller computer for the Intrex display consoles will improve the operating characteristics of the time-sharing environment and may allow frequent feedback with less cost at the central-computer level.

In our sample dialog the user takes the first alternative (statement 7U) and the system responds with the standard output (message 8S) for the matching documents.

At this point in the dialog let us assume that the user already knows how to make an output request or that he refers to his hard-copy version of the Guide. In any case, the user's eighth statement requests additional selected output information. Note that by stringing the "go" command along with the "output" command the user signifies he is sufficiently sure of his statement that he does not want Intrex to respond with its interpretation and timing estimate but rather to "go" on and perform the requested output.

The system then responds (message 9S) as directed. Note the special output information giving those subject terms that matched (MATCHSUB) and the estimated relevance of these terms. Parts 15.74 and 15.75 of the Guide, which are included in Fig. B-4, explain how relevance is estimated for these terms.

In statement 9U the user is selecting, by means of an INFIELD command, a subset of the 7 documents which his original search found. This command enables the user to request only those documents in which a specific character string (here, "harvard") appears in a particular catalog field (here, "affiliation"). Note that, at the same time, the user is changing his output request and using abbreviation "o" for the command name "output" and 71 for the field name "abstract".

In the system's response (message 10S) to the above request, the user has availed himself of the interrupt capability and halted the output at the point indicated by the letters "INT 1" (the additional characters on the next line remain from a program buffer and would not appear in a revised version of the interrupt programs now being debugged). The system then responds with the READY message indicating the user may go ahead with other requests. The interrupt capability is a general facility allowing the user to cut short any system messages. After the user has become familiar with the system he can reduce the verbosity of system messages by entering the SHORT mode. He may do this at any time during the dialog or even upon resuming the system as shown in Fig. B-5.

MONITORING SYSTEM USE

In an experimental system such as Intrex it is important to observe, unobtrusively if possible, the man-machine interactions, the problems that arise, how the user attempts to cope with them, his failures and successes, and so forth. Embedded in the Phase-II system is an elaborate monitoring system which records all console activity, that is, everything which the user types into the system, and everything with which the

```
resume intrex short
W 1355.1
Please sign in.
R
marcus r s/mit 35-406
R
s solid phase transitions/in affiliation harvard/o 22 abstract/go
1. D 3174

(22)
          Lyman Laboratory of Physics Harvard University, >Cambridge<, >Mass.<;/ Lincoln
          Laboratory M.I.T., >Lexington<, >Mass.<

(71)
          The high-temperature series expansion of the zero-field magnetic susceptibility,
          *chi*/*chi**sub Curie* = 1 + *SIGMA**sub l = 1*sup *infinity**a*sub l*(J/kT)*sup l*,
          is related to the diagrammatic representation of the corresponding hh-tINT. 1
 are then
R
s transint##tions/a hoshino/o 21 74 75
S: transit-ions / A: hoshino found: 1 doc o: 21, rel, msub   15
secs/doc.
R
o go
Sorry, I can't understand you.
R

go
1. D 2851

(21)
          Hoshino, Sadao;
          Shimaoka, Kohji (JA);
          Niimura, Nobuo (JA)

(24)
          Ferroelectricity in solid hydrogen halides

  1  docs found
R
quit
Thank you for using Intrex.
R 36.050+10.433
```

Fig. B-5 Sample Dialog in SHORT Mode

system responds, as well as information which the system relates about its internal state as it progresses through a console session. The user is completely unaware of this process and should be relieved of some of the usual problems associated with "over-the-shoulder" observation sessions.

The user has at his disposal a COMMENT command which enables him to make a comment to the system about the system, the cataloging, or the documents in the collection. These comments are, naturally, recorded by the monitoring system. It is anticipated that through this monitoring facility we shall be able to record and observe such things as which system facilities are most effective, what difficulties are encountered in the learning and operational use of the system, which fields within the augmented catalog are most often referenced, and real-time response and computer efficiency (both real time and computer time used at each interaction are recorded). See Fig. B-6 for the first part of a printout of the monitor file created in the sample session illustrated in Fig. B-3.

EXPERIENCE WITH SYSTEM USE

The Demonstration System was operated successfully from West Berlin via telephone-communication lines to the Project MAC CTSS System in Cambridge, Massachusetts during a two-week conference on The Computer and The University held at the Technical University of Berlin (T.U.B.), July 22 - August 2, 1968. Two consoles were set up at the conference to demonstrate the system. One was an IBM 2741 teletypewriter console, the kind employed as standard user I/O equipment for CTSS. The other was an ARDS graphical display and keyboard console of the kind developed at the Electronic Systems Laboratory for Project MAC.* This latter console requires a higher capacity communication link to drive its faster rate of character display. As it turned out, the CTSS time-sharing system and the Intrex retrieval system performed reliably, but a good deal of trouble was experienced maintaining the communication links during the course of the communications. Closed-circuit TV monitoring of the ARDS display in the demonstration room at T.U.B. provided enhanced-viewer presentation to groups of 150 persons.

Controlled experiments have been performed at M.I.T. to test the effectiveness of the user prompting and instructional aids of the Demonstration System with real users. Six users have tried the system so far. These users are staff members of Project Intrex who have not been directly involved with the development of the retrieval system.

---

* For references to ARDS see:

(1) R. H. Stotz and T. B. Cheek, "A Low-Cost Graphic Display for a Computer Time-Sharing Console," M.I.T. Electronic Systems Laboratory Technical Memorandum ESL-TM-316, July, 1967.

(2) R. H. Stotz, "A New Display Terminal," Computer Design, April, 1968.

Fig. B-6 Portion of Monitor File

```
MN0819 277317    08/19/68  1349.9
```

Time file was printed

```
¢014¢014
```

Name of Monitor File: includes date, console no. (277), and indication of time

```
**********************************
time is 17 past 13
probno.  M5806
progno.    160
Greetings! This is Intrex 1a.  Please sign in by typing your name and
address as in the following example:

smith, r j/mit 13-5251

Note that your sign-in statement should end with a carriage return.
TIME ready 4.0/4.0 13.6/13.6
READY
marcus, r s/mit 35-406/
If you already know how to use Intrex, you may go ahead and type
in commands. (Remember, each command ends in a carriage return.)
Otherwise, for information on how to make simple queries of the
catalog, type

info 2

or, to see the Table of Contents (Part 1) of Intrex-1 Guide which will
direct you to other parts of the Guide explaining how to make more detailed
queries, type

info 1
TIME ready 4.0/8.0 38.4/52.0
READY
info 2/
Part 2 of Intrex 1A Guide:  Simple Queries


     To find documents in the system specify your query by subject, author,
and/or title terms, as shown in the 3 examples below:

subject ferroelectric transitions

author Hess, G.B./subject helium

title sulfurization/author Swisher

     In order to specify additional restrictions (e.g., where author comes
from, journal, word variations not to use, etc.), see Part 9 of the Guide
(or type info 9).  For other than standard output (document numbers, titles,
and authors) see Part 8.  For general Intrex command format and abbreviations
see Parts 6.1 and 6.2.
To see Table of Contents for Intrex 1a Guide and how to use
the Guide on line, type


info 1

Otherwise, you may now make simple queries or use any other command.
TIME ready 10.4/18.4 97.8/149.8
READY
subject solid phase transitions/
A search on your query SUBJECT solid phas-e transit-ions found 7 documents.
To output the catalog fields DOCUMENT NUMBER, TITLE, AUTHOR on those
documents type
```

None of them has any specific knowledge of, or interest in, the subject area of the Demonstration data base (spectroscopy of solids and high-temperature metallurgy). Indeed, with only 48 documents this data base is not practical for meaningful retrieval experiments. However, these users were more or less unaware of the specific user-language features of the Demonstration System so that they have provided a good means to test the language and the instructional and prompting features of the system.

In general, the users found the Demonstration System easy to use, though they had specific suggestions for improvements. Learning to use the system was significantly aided by the prompting dialog and the Guide, but it was evidently difficult to learn some aspects of the system without any human help. Users varied in their approach to the system. Some spent considerable time (over an hour) reading the hard-copy version of the Guide before sitting down at the console. Others spent little or no time. Similarly, there were differences in online behavior on such points as extent of use of hard-copy and online versions of the Guide, interrupt facilities, and LONG and SHORT message modes. The experiences so far, while still quite limited, have suggested some particular and general improvements to the system, as discussed below.

SYSTEM IMPROVEMENTS UNDER DEVELOPMENT

A number of system improvements have been planned and are under various stages of development. A partial list is given below.

Programs have been written to display terms in the inverted file and to show how many documents are indexed under each term. These programs have been debugged but not yet integrated into the system. One reason for this is that computer-core-storage space is at a premium. The present programs consume about three-fourths of the 32,000 locations of the core available to programs. The remaining core area is needed for various tables, directories, and buffers. Despite careful control of dynamic-storage allocation, space is still "tight". Therefore, we are planning an overlay system which will enable sections of the system to be read in from disc storage as they are needed.

Work is progressing on improving the efficiency of certain programs. A number of programs were found to take too much time because of the time it takes to open and close disc files. For example, a dialog of the kind given in Fig. B-3 requires about two minutes of computer time.

The experience with users described above has lent increased emphasis to the need for improvements in the dialog which we had previously been planning. These include more complete error messages, more examples in the Guide, and more differentiated dialog for different situations (for example, different numbers of documents found).

One major area under study is the question of matching algorithms and relevance. As indicated above, matching is done on a modified "anding" of all query terms. One would presumably like the user to have the ability to specify any combination of

"ands", "ors", and "nots" among query terms; terms to <u>emphasize</u> within the search specifications, online modifications to the relevance and matching criteria (by term ranges or weights, for example); and selective override to the stemming algorithm. However, all these variations involve complexities in programming and user language, whereas their utility is undetermined, especially in the free-vocabulary indexing situation of Intrex. Therefore, we shall try to get some retrieval experience with the present system before making extensive additions in this area.

Other improvements being planned include: decoding of catalog fields (for example, "English" for "e") on output; more general INFIELD specifications (for example, ranges on dates); more accurate estimates of time to perform output; and naming lists of documents or arguments for later referral by the user.

FREE-FORM INPUT LANGUAGE

The ease of adaptation of the user to the language in which he specifies his requests is recognized as a vital factor in user acceptance of the system. Therefore, a variety of approaches to this problem are being considered. The approach that has been developed most fully is that contained in the Demonstration System, as illustrated in the sample dialog of Fig. B-3. This approach may be characterized as being simple but with a relatively fixed-form language. Another approach characterized by a free-form input language is under study, as described below.

The kind of interaction this free-form input language is intended to make possible is indicated by the following portion of a sample dialog:

S(system) : If, at any time you wish assistance, type a question mark followed by a carriage return. PROCEED.

U (user) silver halides

S : 43 documents found under SUBJECT SILVER HAL-IDES. The following 7 are most heavily devoted to your subject: (prints the title, author, and record numbers of 7 documents)

U : ?

S : At this point, you may:
(1) Obtain more information about these documents.
(2) See more of the relevant documents (if any remain).
(3) Make a new request or alter the given request.
To obtain information about any of these alternatives, type the appropriate number (1, 2 or 3) followed by a carriage return. If these alternatives do not meet your needs or you wish more help, type a question mark followed by a carriage return. PROCEED.

U : 1

S . Among the additional items of information that you may obtain is. . .

Note that in this sample dialog, the user is never required to learn that he should specify that he is looking up a subject term and that to do so he should use the word

subject or the letter s. However, at the point that the dialog is broken off he might begin to learn such details about the system. Even so, unlike the simple online use of the Guide the system would decide for the user which information he is to obtain and he would be told only about field specification.

Although such a system can save the user a certain amount of trouble, it also incurs certain costs to the user. When, for example, the user fails to indicate that he is presenting a subject term, the system might begin by guessing that he has an author in mind. If there is no author with the appropriate name (for example, S. Halides), the only cost is the cost of the additional computer time required to determine that no such author name exists in the system. If, however, by coincidence, an appropriate author name does exist, there is an additional cost in the confusion which may arise. The aim of the guidance-system design is to attempt to minimize such difficulties in the hope that the user who is saved the cost of learning about the system will find the balance in his favor.

The guidance system keeps track of where a user is through a moving point to a set of interconnected nodes, and controls its reactions to what the user says by looking at the contents of the nodes to which the pointer points. Each of the nodes in the interconnected set of nodes (or graph) is a serially ordered set of four-element data structures called subnodes. The first such subnode in a given node controls the operations of the node itself while the remaining subnodes control what is said to the user, what the user may say in return, what is done when he says something, and finally where the pointer is moved to, if an appropriate response is made.

Thus, for example, in the sample dialog segment above, the first message printed to the user is controlled by a node which also interprets his response. Here the user's statement might first be checked (via the second subnode) to see if it is a sign-in; then if this test failed, it would be checked to see if it were a subject term, an author name, and so forth. The reactions are controlled by subsequent elements in the initial node. When a subject-term match is found in the inverted files, control is turned over to the Intrex system, but before this happens, the fourth element in the subnode, according to which the term was recognized, moves the user's pointer to a new node which controls the next response. It is this node that is used to respond to the user's request for more information.

An initial version of this facility is currently being programmed.

DATA-BASE GENERATION

Inputting

The inputting of catalog information stored on paper tape into the system, its online editing, and associated processing has continued in a production routine as described in previous Activity Reports. An experiment to evaluate the implications of online inputting of catalog information via an IBM 2741 directly into the CTSS

(Compatible Time-Sharing System) was performed and results are reported in Section B.1. From a computer viewpoint a significant observation can be made, namely, that the online-inputting procedure appears to be rather sensitive to the host environment (CTSS) and the specific programs which are used to receive the input. Additional investigation is planned when the local Intrex console and buffer/controller become available. There is some hope that this satellite-computer system may make online inputting more attractive.

Programming Changes

As experience was gained in the operation of the programs for Phase-II inverted-file and catalog-record generation, several opportunities for achieving greater programming efficiency were noted. One kind of improvement is the more fully automated chaining of the programs. Another kind of improvement involves the omission of most line-ending characters and tabs as originally typed from the catalog records in order to obtain easier formatting of output on consoles having different maximum line lengths (IBM 2741, about 120 characters; ARDS, about 80; Intrex display console, about 60). The revised formatting program is described below.

As soon as these revised programs are fully debugged we intend to pursue the generation of the inverted file and catalog record data base. We hope to expand quickly the number of documents in storage from the present 48 to at least about 1000 so that we may begin experiments with real users with a data base allowing meaningful retrieval.

Formatting Program

The formatting program takes the edited catalog records from the disc and reformats them for efficient storage and retrieval. The purposes of the program are: to reduce the amount of space required for storage; to store data in a form that minimizes retrieval time; and to detect remaining errors wherever possible.

The formatting program produces three major types of output: an updated catalog-record file produced by adding the new catalog records in its input to the existing file; shred files which provide the raw materials used for updating the inverted files, and the catalog directory file through which the retrieval system accesses the catalog records.

Catalog-Record File. The catalog-record file contains the information about documents that have been entered by the catalogers. Information in this file is structured so that the formatting information (which indicates where records, and fields within records, begin and end) is separated from content information. The formatting information is contained in the header (Fig. B-7) which indicates where a given record or field begins and ends. The rest of the information is stored in the upper and lower bodies. The upper body contains the information that does not require a free format. Such information can be both compressed to save space and preformatted to simplify retrieval. Among the fields included in this portion of the catalog record are the

record number, the level of treatment, language, and the like (see Fig. B-7). For each of the fields included in the upper body there is an escape option for use when the information in these fields does not fit within the pre-established format and must be stored in the lower body.

The bulk of the information in a catalog record is stored as straight text in the lower body. Characters that serve to delimit fields and record entries are removed, as are formatting characters (carriage returns and tabs) whose position is largely a function of the line width of the output device. The average formatted catalog record requires approximately 500 computer words for storage.

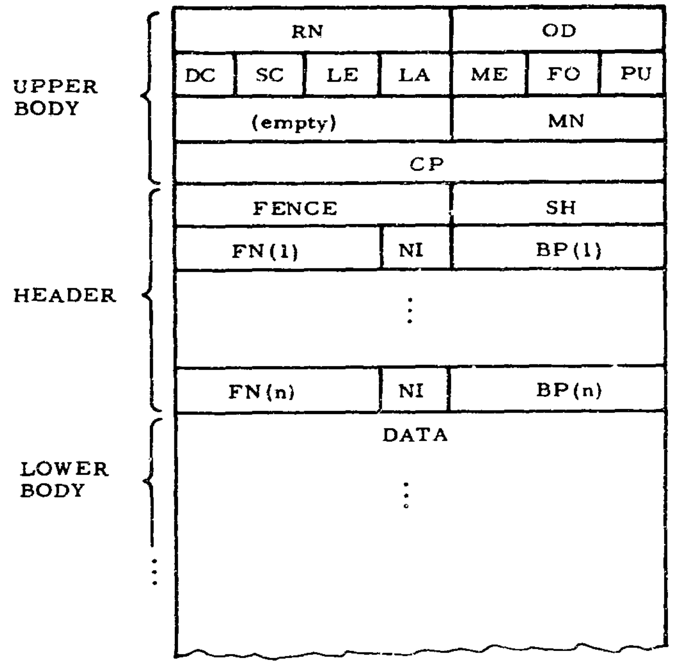
Shred Files. The shred files form the basis for updating the inverted files. Shred files are produced from the information in the author, title, and subject files. Separate shreds are produced for each author last name, and for each word or item in the subject and title fields. These files provide the input for the programs that generate the inverted files by stemming the title and subject words in the shreds, sorting the result, and amalgamating it with the existing inverted files.

The Catalog-Directory File. The retrieval programs access individual records through their record numbers. The purpose of the catalog-directory file is to tie this record number to the physical location of that record within the computer. The first ten words of the catalog-directory file (Fig. B-8) serve to store general information about the catalog records and the directory file itself. The remainder serve to store pointers to the catalog records with the n-th word in this portion pointing to record number n. The catalog-directory file is separated from the catalog records, and all accessing of such records is done via this file, in order to make it possible to change the physical location of a record for purposes of correcting, appending, or otherwise updating records while making a change in only one place in the system.

Program Operation. The formatting program requires approximately two to three seconds to process a single catalog record. Because of the relatively free format of the data within most fields, only a limited amount of error detection is possible and no error correction is attempted. Currently, errors are detected in less than one out of 25 records.

Stemming

The work of Miss Julie Lovins in developing stemming algorithms was mentioned in the Semiannual Activity Report of September 15, 1967. This work, which has been described in a Technical Memorandum (see Section V), includes procedures for stemming words in a two-phase process. In the first phase, the longest possible ending from a list of endings is dropped from the word. Before an ending is dropped it must satisfy a context rule. (For example, do not drop s after s.) The Phase-II Intrex retrieval system includes transformational rules to account for certain spelling anomalies in English (for example, absorb/absorp-tion; split/split-t-ing).



RN: Record Number (Bits 0-20)

OD: Online Date (Bits 21-35)

DC: Descriptive Cataloger (Bits 0-5)

SC : Subject Cataloger (Bits 6-11)

LE: Level of approach (Bits 12-14)

LA: Language (Bits 15-19)

ME: Medium (Bits 20-24)

SH : Size of Header (in computer words)

FO: Format (Bits 25-30)

PU: Purpose (Bits 31-35)

MN: Method Number (Bits 21-35)

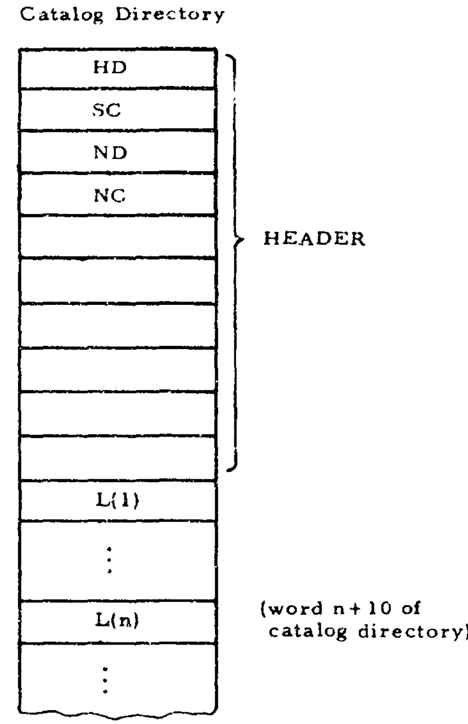
CP: Continuation Pointer

FN(i): Field number of i-th field (address portion)

NI : Note Indicator (tag portion)

BP(i): Byte pointer to bottom of field i (decrement portion)

Fig. B-7 Catalog-Record Format



HD: Highest document number

SC : Size of catalog record file (in computer words)

ND: Number of documents in catalog record file

NC: Number of changes made to catalog record file

L(n): Location of (first word) of record n.

Fig. B-8 Catalog-Directory Format

Because the transformational rules of Phase II involved various complexities in actually performing the stemming and keeping account of it in the inverted files, and because the number of additional cases it handled seemed relatively small it was decided to try out only the Phase-I procedures in initial Intrex systems. A list of endings and context rules for applying them are contained in the Lovins memorandum.

Several thousand subject-term words from several dozen catalog records have been stemmed according to these Phase-I procedures. So far the results seem very good and only a few modifications to the original procedures appear desirable. A few endings have been added and the minimum stem length was raised from two to three letters. Figure B-9 gives a statistical compilation of the number of endings found for each ending type in one run of 2,382 words stemmed.

MESSAGE-FILE MAINTENANCE

We are well aware that feedback is essential for effective interaction, but we also recognize that the experienced user may lose patience with the verbose messages which are appropriate for the less experienced user. With an eye to this problem, our output routines have been designed and implemented to print symbolically referenced, externally stored (on the disc) messages interposed with in-core parameters. By keeping references to messages symbolic (for example, as BCD message names), we can easily control what a particular message says. This level of control is useful to modify message content as the dial evolves as well as controlling the verbosity of the dialog. The current system permits only a binary choice, LONG or SHORT. Whenever a command is given to change the mode, the name of the message file (the message files are stored on the disc) is changed. This in turn causes the messages to emerge in their full or abbreviated form. In the future, it may be advantageous to provide one or more intermediate levels of verbosity between the extremes of LONG and SHORT.

The Guide and other standard dialog messages are kept in disc files except for a few common short messages which are kept in core. Messages may be readily changed by standard CTSS editing programs. The maintenance of the Guide, described below, is illustrative of the required procedures.

Two versions of the Guide are required, one for the hard-copy edition, the other for online use. The hard-copy Guide must be generated by a file containing instructions for the manuscript format generating program, ROFF. The online version must contain message labels and delimiters for response to online requests about INTREX operation.

To facilitate updating of the Guide, an ASCII master file is kept. This file contains both the online version labels and delimiters and the ROFF control instructions. Corrections and revisions are made on this master copy through use of the CTSS-editor program designated QED. If an online Guide version is desired, one first issues approximately five QED requests to delete the ROFF instructions, and

Ending	Occurrences	Ending	Occurrences
arization	9	ness	9
entations	2	ogen	20
ableness	1	wise	3
entation	3	ying	1
ability	2	age	4
ational	5	als	9
ibility	16	ant	7
ization	23	ary	5
ations	20	ate	15
encies	3	ely	1
ential	3	ene	1
istics	2	ent	30
acity	10	ial	8
aries	1	ian	8
arity	4	ics	10
ately	1	ied	4
ating	25	ier	3
ation	105	ies	16
ative	3	ine	14
ators	2	ing	83
atory	1	ion	215
ement	11	ism	4
ening	3	ity	30
ental	2	ive	5
ially	1	one	1
icity	6	ons	3
ional	10	ors	7
istic	1	ous	3
ities	4	's	24
ivity	8	al	57
able	9	ar	27
ally	1	ed	106
ance	24	en	34
ants	16	es	148
ated	12	ia	1
atic	2	ic	127
ator	7	is	12
ence	31	ly	2
ency	15	on	11
ents	30	or	11
eous	2	um	30
ible	4	a	25
ical	62	e	449
ions	40	i	23
ious	1	o	16
ized	11	s	118
less	3	y	85

Fig. B-9 Endings Found in Stemming 2,382 Words

then one processes this Guide through the message directory generator, DIRGEN. For a hard-copy version, one uses QED to edit the labels and delimiters of the master copy of the Guide and processes the resulting file through ROFF. The ROFF output file may then be printed at the console or saved, written on magnetic tape, and then processed through the Project MAC 1401 computer with high-speed printer.

The present hard-copy edition (24 July 1968) was produced on ditto masters at the console; the ditto masters were used to make copies of the 40-page guide.

AUTOMATIC INDEXING

New methods for generating subject indices automatically have been studied by William R. Kampe as part of his work toward a Master's degree in Electrical Engineering. The results of this work are reported in Mr. Kampe's Master's thesis which has now been issued in slightly revised form, as a Technical Memorandum. (See Section V.).

The main motive behind this research was to develop techniques for getting documents into the augmented catalog more rapidly. Since it now appears that more than half the time required to catalog a document is devoted to the subject indexing, the use of techniques that could, at least temporarily, permit the omission of this step, could result in a considerable decrease in the amount of time required to enter a document into the data base.

All the techniques studied derive the subject words from the title and the abstract of e document. This information is included among the descriptive-cataloging information that has to be entered anyway, so that no additional inputting is required. An empirical comparison of the words appearing in the subject-index field with the words appearing in these two fields showed that over ninety percent of the words appearing in the subject field also appeared in the title and/or abstract fields so that the automatic generation of subject words from these fields seemed possible, at least in principle. Furthermore, Salton* has found that the indices generated from the title and abstracts of journal articles tended to be approximately as good as indices generated from full text.

Three methods for pre-indexing documents were compared. The major distinctive features of these techniques is their reliance on previous human indexing for deciding which words are relevant and the use of this information to discard, rather than retain, words for indexing. Comparison of the efficacy of these techniques was made by matching their results with those obtained, independently, by human indexers.

---

* Salton, G. (ed), "Information Storage and Retrieval," ISR-11, Department of Computer Sciences, Cornell University, Ithaca, N. Y., June, 1966.

All three techniques are based on the "usage rate" of terms in previously indexed documents. This usage rate is a single parameter that indicates how often the given word has been used in the subject index when it appeared in the title or abstract of the document. In pre-indexing Method I, words are retained if their usage rate exceeds a certain threshold. (The efficacy of the method depends on the threshold, and several levels were tested.) In Method II words with high usage rates are also discarded on the grounds that they are either function words ("the", "of", "in" and the like) or very common words in the collection (such as "magnetic") neither of which convey much information about the document. Again, of course, this method depends on the "commonness" threshold, and again, several levels of the threshold were tested. Method III is like Method II except that an ambiguous area is neither very high nor very low. Decisions in this ambiguous middle area are made on the basis of the "context". If either of the neighboring words is in the area of unambiguous inclusion then the given word is included, otherwise it is not.

Evaluation of these methods was made on the basis of two criteria: completeness, or the percentage of appropriate words that could have been included and were, and relevance or the percentage of inappropriate words that were excluded. Appropriateness, in both cases, was determined by seeing if human indexers had selected the words.

When this criterion was taken without interpretation, Method I was the most successful, achieving over 90-percent completeness with over 60 percent relevance at its optimal thresholds. However, much of this success was seen to be based on the inclusion of function words and other words with little information content. When these were excluded, Methods II and III were seen to be better with Method III attaining both relevance and completeness ratings of over 90 percent at its optimal setting. These results suggest that rapid pre-indexing along the lines investigated is feasible.

AUGMENTED CATALOG STORAGE DEVICE

We are completing specifications for an experimental test setup for the purpose of demonstrating feasibility and predicting performance of the magnetic-tape device described in the preceding Activity Report. The most radical departure in the approach under investigation from current tape transports occurs in the utilization of the magnetic-transfer process from the tape to the cylinder. Accordingly, the experiment will concentrate on maximizing bandwidth of the transfer and read processes. This involves not only linear information density, but also the tape velocity.

In order to facilitate experiments, the test facility will differ from the proposed device previously described. Tape reels are eliminated and a continuous loop of tape is utilized. We desire to be able to both read and write on the same device.

This is accomplished by providing a write head which operates in contact with the tape at writing velocity (less than 100 inches per second) and over which the tape floats at reading velocity (1,000 inches per second).

Our experiment imposes two constraints on the magnetic cylinder assembly. First, we desire to try several different thicknesses and types of coatings, and thus would like to be able to replace the cylinder easily. Secondly, the runout of the cylinder must be carefully controlled so as to permit in-contact reading at high velocity. These constraints will be met by use of a precision grinding spindle. Each cylinder blank will be mounted in its own wheel holder, ground true on the same spindle, and subsequently plated. It can then be easily and accurately mounted on the spindle when needed.

Such an experimental setup will demonstrate the writing, transfer, and reading of digital data under conditions similar to those of the proposed device. It will allow us to determine how variations in coating thicknesses, materials, and writing and reading techniques affect the bandwidth at which data can be read reliably.

3. THE DISPLAY-CONSOLE SYSTEM

Staff Members

Dr. D. R. Haring
Mr. J. E. Kehr
Professor J. K. Roberge

Graduate Student

Mr. P. F. McKenzie

Undergraduate Student

Mr. A. Spencer

SUMMARY

In preceding Semiannual Activity Reports a complete system design for the augmented-catalog display-console system and the functional block diagram of the display console itself have been described. During the present reporting period the construction of the electronics and logic for the console has been completed and they are presently undergoing tests. The logic design of the buffer/controller has also been completed and is presently being constructed. Minor changes have been made in the system configuration described in the preceding reports in order to improve system performance and to take advantage of special features of the detailed properties of the logic building blocks being used.

The character generator described in the preceding Semiannual Activity Report is now operating at the proper speed and produces high-quality characters. It is presently being integrated into the logic of the console.

Work continues on the 620I buffer/controller software, some of which is described here. Current experimental investigations being conducted on the display-console system are also presented.

SYSTEMS CONSIDERATIONS

The buffer/controller (B/C) of the display-console system serves as the interface among several augmented-catalog consoles, the text-access central station and the time-shared computer. The major elements contained in the B/C are a small processor to manage the data flow and a serial memory to refresh the CRT displays of the consoles. The serial memory also provides back-up memory for the small processor. Various features of the B/C have been discussed in previous activity reports. Details of its design and operation not previously reported are presented in the paragraphs that follow. A functional block diagram is shown in Fig. B-10.

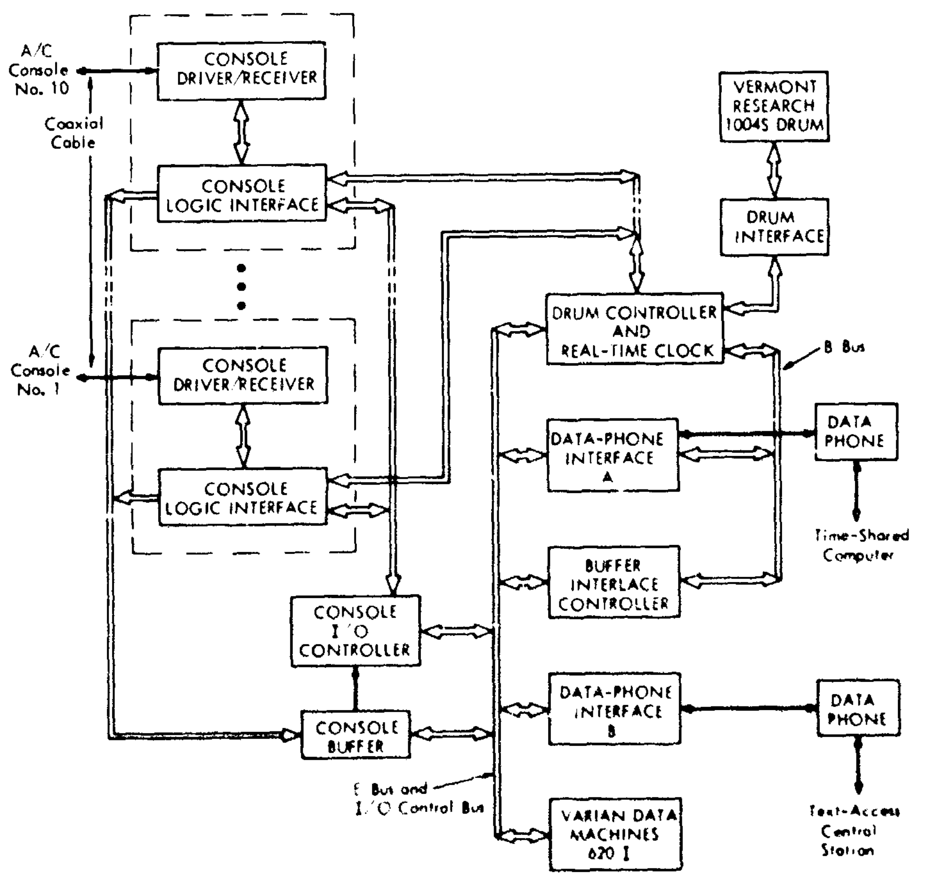


Fig. B-10 Function Block Diagram of the Buffer/Controller

Project Intrex is designing and constructing the logic that is contained in the B/C. The purpose of this logic is to interconnect the serial memory and small processor with the various external devices. The logic design of the B/C has now been completed and the logic is presently being constructed.

The Varian Data Machines 620I manages the data flow in the B/C through connections to its E Bus and input/output (I/O) Control Bus. When the B/C is first turned on and after the Vermont Research 1004S drum has attained operating speed (approximately four minutes) the SYSTEMS RESET button is actuated on the 620I control panel. This places the B/C at a known initial state. The 620I operating programs are then loaded and the system is ready for operation.

To provide a foundation for our discussion about the operation of the B/C, let us first briefly describe the operational properties of the 620I computer. The standard 620I communicates with peripheral equipment on the 16-bit word parallel (I/O) E Bus and 12-line I/O Control Bus. Each information transfer occurs under control of a stored program. Information exchanges with peripheral devices are synchronized by their controllers. Each controller and the device that it controls comprise a peripheral device. The basic machine provides four types of I/O operations:

External Control:	An external control code is transmitted under program control from the computer to a peripheral controller.
Program Sense:	The status of a selected peripheral controller is interrogated by the computer under program control.
Single-Word Input Transfer:	A single 16-bit word of data is transferred under program control from a peripheral controller to the 620I A register, B register, or any memory location.
Single-Word Output Transfer:	A single 16-bit word of data is transferred under program control to any peripheral controller from the 620I A register, B register, or any memory location.

In addition, the 620I used in the B/C has a Direct-Memory-Access-and-Interrupt (DMA/I) option. With the DMA/I option the computer is capable of I/O data transfers to or from memory without program intervention. This option allows external devices on the I/O bus to request two types of functions. First, the interrupt feature of the DMA/I option permits an external device to request the INTERRUPT FUNCTION. When such a request is made, an instruction, which is independent of the main program, can be executed. When two or more requests are received simultaneously, a hard-wired external-device-priority scheme determines which device has priority. The computer is directed to a memory location specified by the interrupting device and is caused to execute the instruction found at that location. Any 620I instruction, other than an I/O command, may be executed in this manner. Normally, however, the instruction will be a jump-and-mark

instruction and will result in the processing of an I/O subroutine. Secondly, the interrupt feature of the DMA/I option also permits an external device to request the TRAP FUNCTION. When such a request is made, the trapping-device controller can either request that the 620I output one word of data from memory (trap out) or request that the 620I input one word of data to memory (trap in). The address of this word is placed on the E Bus by the controller after the 620I acknowledges the trap request. The stored program is caused to remain idle for 2.7 microseconds during which time the transfer takes place. This "cycle-stealing" action does not disturb the operational registers of the 620I, thus allowing the stored program to proceed normally at the conclusion of the data transfer. The trap function is also connected to the priority logic.

When the B/C is completed, it will also have a device that is equivalent to the Buffer-Interlace-Controller (BIC) option which is available for the 620I. The BIC generates the addressing and timing signals that control data-word transfer between the selected peripheral device and the computer through the trap function. This option controls the transfer of data blocks between the memory of the 620I and peripheral I/O devices concurrently with program operation in the 620I. Full 16-bit word transfers can be controlled at rates up to 202,000 words per second (3.232 megabits per second).

Further details about the operation of the 620I and its options can be found in the appropriate Varian Data Machines manuals.

Observe in Fig. B-10 that six peripheral devices are connected to the E Bus and I/O Control Bus of the 620I. Each of these devices has at least one address by which it can be identified by the computer. Only the Drum Controller and Real-Time Clock, the Data-Phone Interface A, and the Buffer Interlace Controller (BIC) have interrupt capabilities. The Drum Controller and Real-Time Clock has the highest priority, followed by the BIC Buffer Interlace Controller which, in turn, is followed by the Data-Phone Interface A. The remaining peripherals operate strictly under program control by the computer.

When commanded to do so by means of the B Bus, the BIC can control data transfers between the 620I and either the Data-Phone Interface A or the Drum Controller and Real-Time Clock. To increase system flexibility the Drum Controller can be commanded to perform much like the BIC to transfer fixed-length blocks of data. The block length is equivalent to one line of data that can be displayed on the console CRT (56 characters).

External-Control commands allow the 620I to control the various modes of operation of the Drum Controller, and Program-Sense commands allow the 620I to determine the status of the data registers in the controller. Finally, provisions have been made to connect the controller to the BIC in order to do arbitrary-sized block transfers of data.

The Real-Time Clock is used to time events in the console system. It is simply a 16-bit binary counter that is advanced by one for each drum revolution. Since the drum turns approximately 57.5 revolutions per second, a new epoch occurs approximately every 20 minutes. The computer can request the reading of this clock at any time by a DATA IN command.

In the initial system, communications between the time-shared computer and the B/C is by means of a data phone, which is connected to the 620I by means of Data-Phone Interface A. This interface is designed to operate through a standard 1200-bps full-duplex data phone and, with minor modifications, to operate up at the full bandwidth of the 620I. In addition to the standard DATA IN and DATA OUT commands, data transfers to and from the time-shared computer can operate on the interrupt function of the 620I.

By External-Control commands it is possible to mask out input and output interrupts selectively. Program-Sense commands allow the 620I to determine the status of the data registers in the interface. Finally, provisions have been made to connect the interface to the BIC to block-transfer data.

Communications between the B/C and the text-access central station is by means of a data phone, which is connected to the 620I by means of Data-Phone Interface B. This interface is designed to operate through a standard 150-to 300-bps full-duplex data phone. Only the standard DATA IN and DATA OUT commands can be used to transfer data between the 620I and the text-access central station. Furthermore, data are permitted to flow in only one direction at a time.

Program-Sense commands allow the 620I to determine the status of the data register in the interface.

The augmented-catalog consoles are connected to the 620I I/O buses by means of the Console Buffer and the Console I/O Controller. The former buffers the data that are transferred from any console to the B/C by means of two 16-bit registers. Program-Sense commands allow the computer to determine the status of these data registers, and DATA OUT commands transfer the data-register contents to the 620I.

A register that is composed of a single flip flop in each Console Logic interface, but addressed through the Console I/O controller, stores the "request-for-service" signals from the consoles. Upon execution of the proper DATA IN command, this register is transferred to the 620I.

A second register that is made up of a single flip flop in each Console Logic Interface, but addressed through the Console I/O controller, is set by the proper DATA OUT command to initiate the transfer of data from a selected console to the Console Buffer.

Figure B-10 shows that data are sent directly from the drum to the augmented-catalog consoles without going through the E Bus of the 620I. The role of the 620I in this case is simply to select which drum track is being viewed by what console.

A change has been made in the method of sending the INTERROGATE signal from the B/C to an augmented-catalog console. Previously, it was reported that this signal would be a burst of sine waves. Because it is simpler from the system's viewpoint, the INTERROGATE signal is sent by making one of the data pulses from the B/C to the given console longer in time duration than all other data pulses.

Detailed functional block diagrams of the individual elements contained in the B/C will be discussed in future activity reports.

CHARACTER-GENERATOR AND DISPLAY-TUBE CIRCUITRY

The operation of the flying-spot scanner character generator has been described in earlier reports. The initial display system, consisting of this character generator and an entertainment-quality cathode-ray tube has been completed and this system is operational. Figure B-11 illustrates a display produced by the

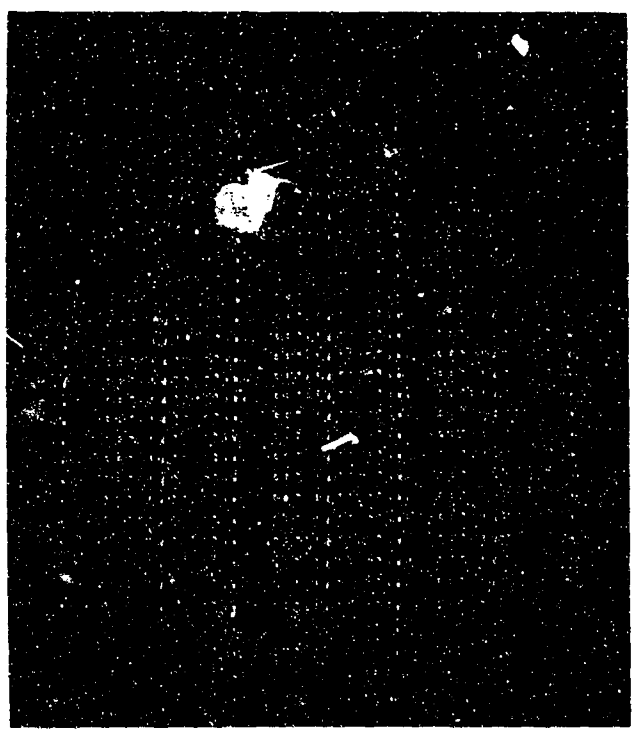


Fig. B-11 A Display of the Output of the Character Generator

system. The full 56 by 31 character display is indicated in the figure. The display is refreshed at a nominal 60-Hz rate.

The current equipment demonstrates the feasibility of the proposed character generation and display system. Only minor changes to further improve character

quality are required to convert this prototype system to final form. The cathode-ray tube will be changed to 14-in. diagonal measure from the 16-in. unit now used. The new tube will have a 90° deflection angle rather than the 110° angle of the present tube, and will incorporate improved electron optics. These changes are required because of the beam defocusing which occurs at large deflection angles with the present tube. Several tubes are now being evaluated. A higher quality yoke, available from Celco Electronics, will be used to deflect the beam of the new tube. These modifications should reduce the spot size by approximately a factor of two compared with the present system, and produce characters over the entire screen with better quality than those now obtained in the center of the display.

The beam-positioning and other required circuitry for a monoscope character generator have been designed and will be tested in the near future. It will be possible to use the monoscope generator as a direct replacement for the flying-spot-scanner character generator in situations where the larger capacity (256 characters for our display compared with 96 for the monoscope) and ease of alteration of the character set are not required. The monoscope generator is physically smaller and less expensive than our flying-spot-scanner unit.

A modified display system which may result in the design of a single console suited to both catalog and text-access use is also under investigation. As explained in the next section, one of the text-access stations will use a Tektronix Type 611 Storage Display Unit to provide a transient display of full text. This display unit has sufficient resolution for catalog use, but has several disadvantages when operated in the storage mode. The inflexibility of the stored display compared with a refreshed display prevents the use of a light pen and increases the demands on the central processor when editing is required. Furthermore, the lifetime of the tube in the storage mode is relatively restricted, leading to high operating costs.

It is possible to use the storage tube in a non-store mode, and in this mode the disadvantages indicated above are not present. It therefore seems that the use of the single display tube in a storage mode for text access and in a non-store, refreshed mode for catalog use may result in a single, economical console which combines these two major functions.

The deflection bandwidth of the Type 611 Storage Display Unit is insufficient for a 2000-character display refreshed at a 60-Hz rate. However, the techniques which have already been developed for the existing character-generator and display-tube system can be used with the Tektronix display. A Type 611 Display Unit has been ordered for evaluation in this application. The unit will be modified by including the sinusoidal scan signal and by increasing the bandwidth of its blanking circuitry. The modified display unit will then be evaluated using the existing character generator.

C. THE TEXT-ACCESS PROGRAM

Staff Members	Graduate Student
Dr. U. F. Gronemann	Mr. A. Jagodnik
Dr. D. R. Haring	
Mr. D. R. Knudson	**Undergraduate Student**
Mr. . E. Kehr	
Professor J. K. Roberge	Paul Kudirka
Mr. P. R. Scott	
Mr. S. N. Teicher	

1. SUMMARY

The design for the experimental text-access system is essentially complete and the system is being assembled. The basic features of the system are described in the preceding semiannual report; additional details are described below. The system will provide remote access to the full text of the approximately 10,000 documents being included in the augmented catalog. Full text will be stored on microfiche. The text-access system will be used in information-retrieval experiments in conjunction with the catalog. Two types of user terminals are included in the initial system. One terminal provides 35-mm film facsimiles of requested text, and the other utilizes an erasable, electronic storage-tube display.

The camera-and-processor equipment for the film terminal has been completed and the film-cutting and transport mechanisms have been tested successfully. The logic circuits for controlling this terminal are currently undergoing tests.

The modified Houston/Fearless Compact Automatic Retrieval Device (CARD) unit, which we shall use for microfiche storage and retrieval, has been received and is operating properly after correction of misalignments which occurred during shipping. The flying-spot scanner is mounted on the CARD unit and the integrated operation of the scanner with the retrieval device will begin shortly.

A working model of the video-signal transmission subsystem has been completed and tested. This equipment is now awaiting integration into the rest of the text-access facility and system testing with the central-station logic circuits. An analysis of various transmission-network configurations based on loading, distance, frequency characteristics, and so forth is being formulated.

The design of the text-access-system control electronics is complete and circuit boards are being fabricated. The Tektronix Type-611 Storage Display Unit for the transient-display terminal has been received and the circuits for controlling the terminal are being assembled.

It is expected that initial operation of the experimental text-access system will begin early in the next report period.

2. SYSTEM DESIGN

Figure C-1 illustrates the configuration of the experimental text-access system. The central station includes a modified Houston/Fearless CARD unit which

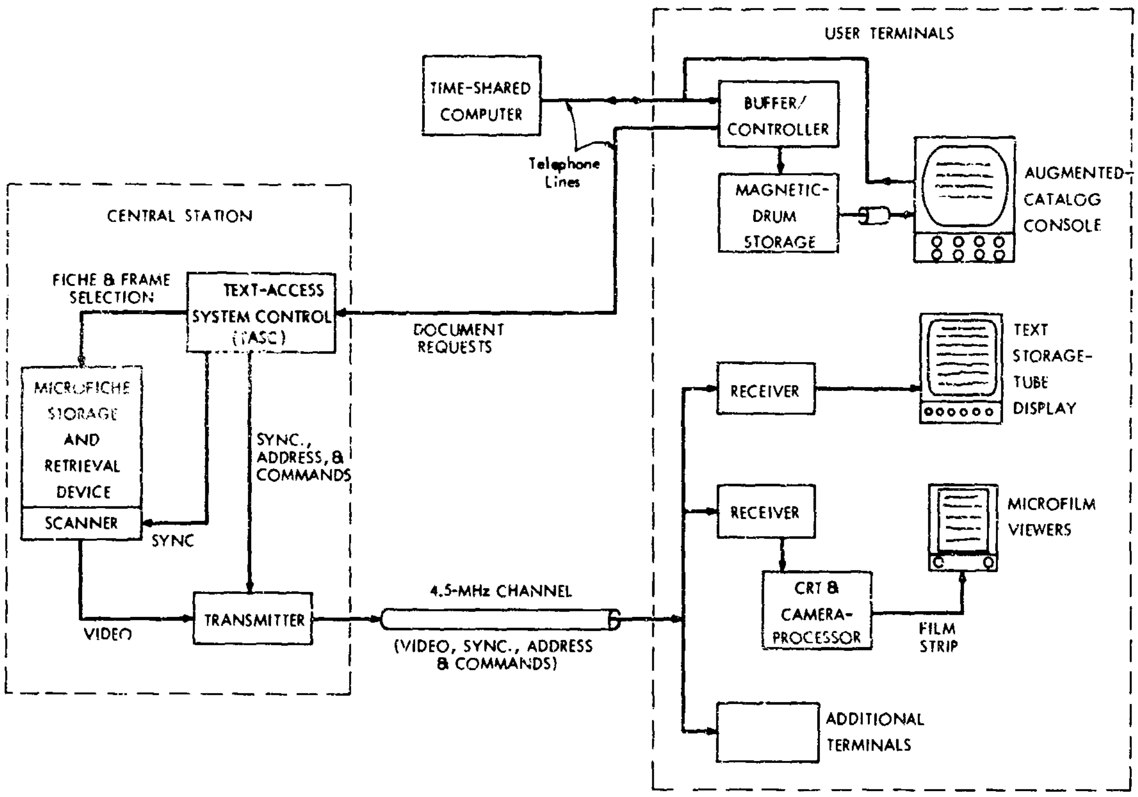


Fig. C-1 Experimental Text-Access System

stores and retrieves the microfiche containing the text images, a flying-spot scanner for converting these images to a video signal, the text-access-system control electronics (TASC), and the transmitting electronics. The retrieval device can access any microfiche within five seconds and can change to any page on a given microfiche within one second. The transmitter combines the video with appropriate synchronizing pulses and coded pulses for addressing and controlling the receiving terminals. A page of text is scanned only once per request and is transmitted over a 4.5 MHz-bandwidth transmission network. The receiving terminals decode the address pulses and receive only the video signals directed to that terminal.

Two types of terminals are included in the initial system, a film terminal producing 35-mm film images of the scanned text and a transient display utilizing a Tektronix-611 Storage Display Unit. The film terminal reconstructs the text

image from the video signal on a high-resolution CRT and records the single-scan image on film. Transmissions to this terminal require approximately 1 sec per page and approximately 1 sec between pages. Multiple pages are recorded on a film strip with a maximum of 10 pages per strip. An automatic processor coupled to the camera delivers the dry, developed film strips one minute after the last transmission. This type of terminal is the only available type which can achieve adequate resolution for storing text from professional-journal-type documents, and it provides a form of remote hard copy for off-line reading.

The transient display is co-located with the augmented-catalog console and provides a quick-look capability for scanning the text of documents associated with searches in the augmented catalog.

The limited resolution and writing rate of the storage tube necessitate system-parameter modification when this unit is used. Accordingly, a 1000-line scan and a full-page writing time of four seconds are used with the storage tube. In an effort to partially overcome the limited resolution of the storage tube, we are providing external circuitry which will allow the user to request a factor-of-two enlargement of a quarter-page centered on one-of-nine selectable points of a page.

Requests for documents can be placed either through the augmented-catalog console being developed by INTREX or through the teletypewriters used with the central time-shared computer. The augmented-catalog console will contain a text-access mode during which the programmable buttons and keyboard are utilized to input user actions associated with text-access requests. Function switches at the text-access display will be provided to flip microfiche frames of a document in either direction.

Some effort has been made in the design of the central-station electronics to minimize the load imposed by the text-access function on the augmented-catalog buffer/controller. This minimization of load is accomplished by implementing many of the control routines in the central-station hardware rather than through computer software.

3. STORAGE AND RETRIEVAL DEVICE

The Houston/Fearless CARD device was delivered in June of this reporting period, but was damaged in transit. The alignment of the selection and retrieval mechanism was upset, and this upset caused very poor reliability during the retrieval cycle. Retrieval malfunctions, in turn, resulted in damage to the set of microfiche used for testing the unit. Since realignment, the device has functioned well and it is expected that, upon procurement of a new set of microfiche, its performance should be satisfactory.

The CARD device is designed to store up to 750 microfiche with access times of less than five seconds to any microfiche and one second to any frame on that

fiche. The microfiche conform to the COSATI standard and contain five rows by 12 columns of page images. The storage capacity of 45,000 pages is expected to be adequate for most of the 10,000 documents being selected for the initial augmented-catalog system, since these documents are averaging approximately five pages per document. The microfiche are stored in this device in bins located on a carousel with 20 to 26 microfiche per bin. The maximum of 26 fiche per bin provides storage approximately 10 percent above the design capacity of 750 microfiche.

The modifications which have been made to the standard CARD unit are those required to integrate the device with a flying-spot scanner and to control the fiche- and frame-selection circuits from computer-generated inputs. The scanner is the same as that used for the experimental image-transmission system described in previous reports. The CRT of the scanner and the CARD unit are mounted on the same base to maintain rigidly the distances between the scanner CRT, lens, and the film plane. A photograph of the modified CARD unit and scanner is shown in Fig. C-2.

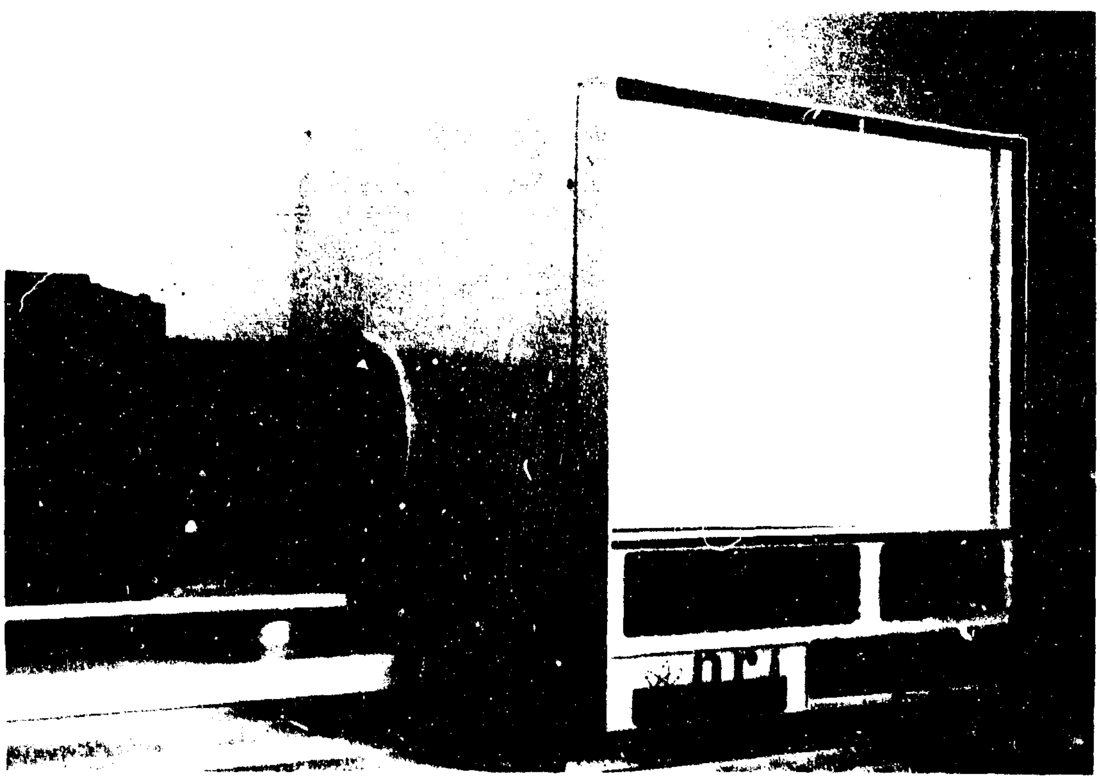


Fig. C-2 The Modified Microfiche Storage Unit and Flying-Spot Scanner

The projection lamp used for viewing in the standard CARD unit is replaced with the photomultiplier tube for the flying-spot scanner. For focusing purposes, the scanner lens can be moved along the optical axis by means of an electric motor. The electrical modifications to the standard unit permit the retrieval cycle to be controlled by signal voltages from the central-station logic circuits in addition to the standard CARD keyboard. The interface between the central station and the CARD unit is described in Section 7.

4. TRANSMISSION SUBSYSTEM

The development effort during the reporting period has resulted in a working model of the transmission subsystem described in the preceding Activity Report. Two important changes in the system have been made: a programmable line counter had been added to the transmitter, and a decision has been made to employ baseband transmission exclusively in the first experimental text-access system.

A diagram of the transmission subsystem in its present form is shown in Fig. C-3, where, for simplicity, only one receiver is shown. Scanning parameters, such as horizontal period and number of lines, are digitally programmable since they vary from terminal to terminal. The programmable counters and crystal oscillator used to implement this feature result in scanning parameters which are accurate to within one part in 10^7. The digital signals will originate in the Text-Access-System Control logic (TASC logic) described in another section of this report. The address and command signals mentioned in the last progress report are also generated by the TASC logic.

The line driver has been designed and built as a standard circuit which will also serve as a video amplifier in certain user terminals. This amplifier combines the digital and analog signals and drives a 75-ohm coaxial cable with a 36-volt, peak-to-peak, signal. The frequency characteristics of the cable can be compensated by means of suitable feedback networks in the line driver.

The receiver analog circuitry shown in Fig. C-3 consists of a variable attenuator, voltage follower, peak detector, and comparator connected as an automatic-gain-control (AGC) system, and a matched-filter and threshold comparator which serve to separate the digital signals. The AGC system compensates for 10-to-1 variations in the low-frequency gain of the system which might result from temperature changes or changes in separation distances between transmitter and receiver. The receiver digital circuitry was discussed in the preceding Activity Report.

A set of rules for the formation of a network of receiving terminals is under development. Based on loading, frequency characteristics, distance, and other considerations, these rules will determine the extent and form of text-access systems which use the standard components. It is evident that the maximum distance

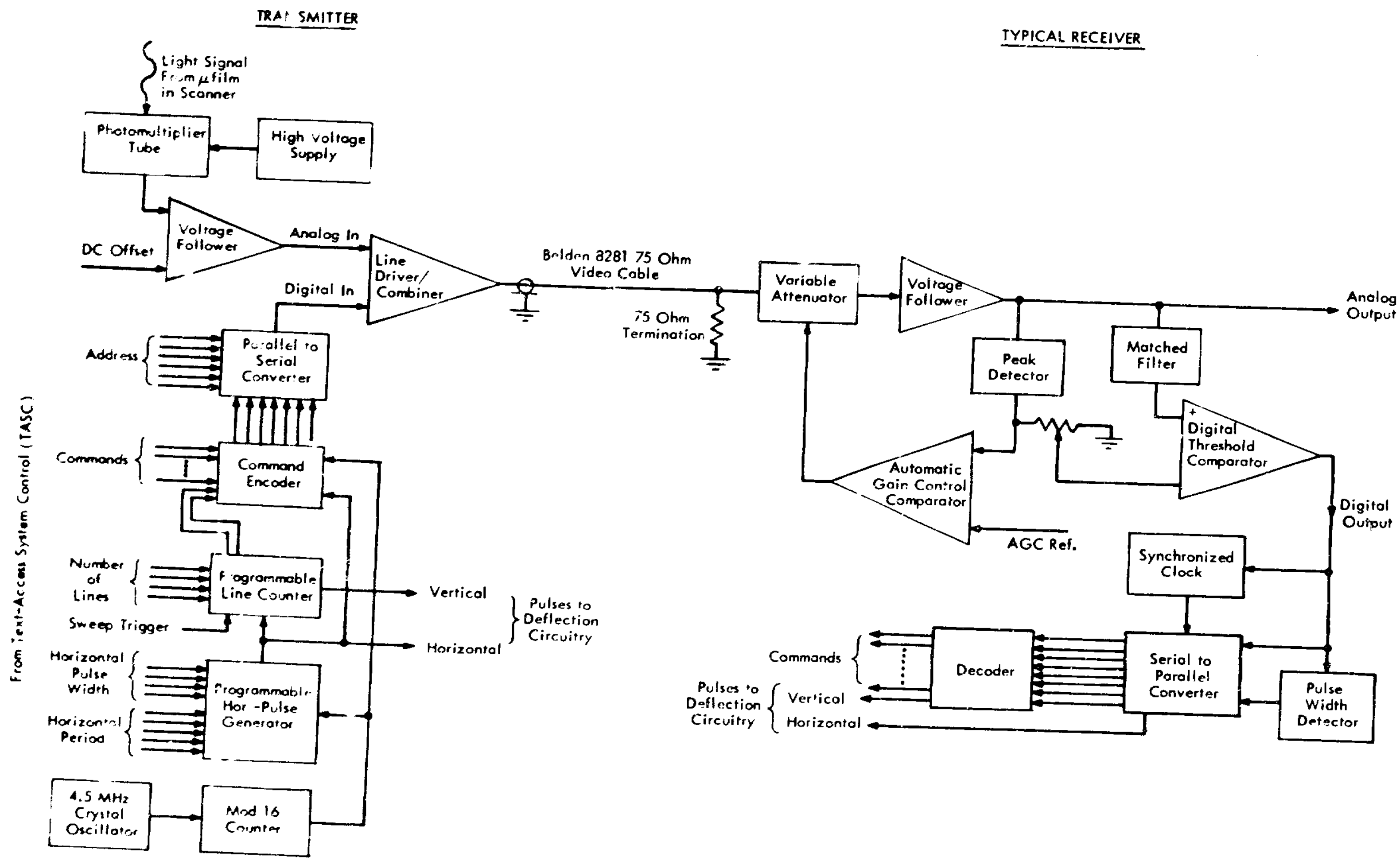


Fig. C-3 Block Diagram of the Text-Access Transmission

will be determined by the transmission subsystem while the maximum number of terminals will be determined by limitations of the scanner and automatic-retrieval mechanism.

Because of the experimental nature of the text-access system, an attempt has been made to design the transmission subsystem flexibly and with an expansion capability. These features are reflected in the following facts about the system:

> The programmable digital circuitry in the transmitter is capable of generating a wide range of sweep rates and numbers of lines.
>
> Provision has been made for six address and seven command bits, although only two addresses and twelve commands are required for the first text-access system.
>
> The crystal-controlled blocks and the AGC circuitry minimize the number of adjustments required; receiver terminals can be moved at will without any design changes.
>
> If carrier transmission is found necessary, the signal format need not be changed.
>
> The choice of 4.5 MHz for the basic clock frequency permits the addition of a pilot carrier to eliminate the crystal oscillators now used in the receivers. This pilot carrier could also be used to provide a high-frequency-gain reference for adaptive long-distance-cable compensation.

The system is versatile, yet not excessively complex. Future terminals can be added as required without modification of transmission-subsystem elements.

5. TRANSIENT-DISPLAY TERMINAL

The design and construction of the transient-display terminal utilizing an electronic storage tube is nearing completion. The terminal will be used in conjunction with the augmented-catalog console and will be located adjacent to the console. Requests for full text of documents are initiated through the augmented-catalog console operating in the text-access mode by means of the programmable buttons and function switches. The programmable-button labels assist the user in selecting various options during the text-request sequence such as the choice of document and page number, choice of storage-tube or film-terminal display, a full-page display or a magnified display of a page sector, and so forth.

In addition to these input devices at the catalog console, three function switches are located at the storage-tube display and connected to the console. Two of these, page forward and page backward, enable the user to request the following or preceding page by pushing a single button. The third function switch initiates the

magnify mode in which approximately one-quarter of the page area is scanned and transmitted, thereby giving a factor-of-two magnification of that page sector. In this mode, an illuminated rectangle appears as an overview on the page display. The rectangle outlines the page sector to be magnified and can be moved to any one of nine positions by means of a pushbutton matrix at the display.

Several indicator lights are included at the storage-tube display to inform the user of the status of his request such as FICHE NOT FOUND, LAST PAGE OF DOCUMENT, and REQUEST IN PROCESS. The pushbuttons and indicators are lighted and operated in a programmed sequence to assist the infrequent user in his requests.

The terminal equipment consists of two main components, the Tektronix Type-611 Storage Display Unit and the electronics for controlling the display and the user inputs. The terminal is designed to be self-sufficient in that it requires no external power supplies. The storage tube has provision for erase, and x, y, and z axes inputs. A dual-channel ramp generator, synchronized to the scanner by the receiver electronics, provides the x- and y-deflection signals. The video signal is applied to the z axis to unblank the tube and to store the text display. Since the storage tube is a bistable device, it does not have gray scale capability.

The rectangle displayed during the magnify mode is generated by appropriate deflection signals to the x and y axes and is refreshed 60 times per second to eliminate flicker. The rectangle is not stored by the display in order that it can be moved to different sectors by the user. The write-through mode for the rectangle is achieved by applying a one-microsecond pulse to unblank the z axis at a repetition rate of approximately 20 KHz. Because the dot writing time of the tube is approximately 20 microsec, the rectangle is visible but is not stored. Although the storage-tube resolution is not sufficient for high-quality textual images, this terminal should permit useful experiments with an erasable, stored display. The magnified-display capability should enhance the usefulness of the device, particularly with documents containing small characters.

6. FILM TERMINAL

As reported in the preceding Semiannual Activity Report, we have purchased a Kodak MCD-II film unit and a GAF Transflo 1206 processor and have merged the two devices into a camera-processor unit. The camera-processor unit, shown in Fig. C-4, is in the final stage of adjustment.

Figure C-5 shows the two electrically-activated magnetic-clutches which couple the transport drive-rollers to the camera motor during film advance and to the processor motor after the film has been cut. Three microswitches are located in the film guides to sense the travel of the exposed film through the transport.

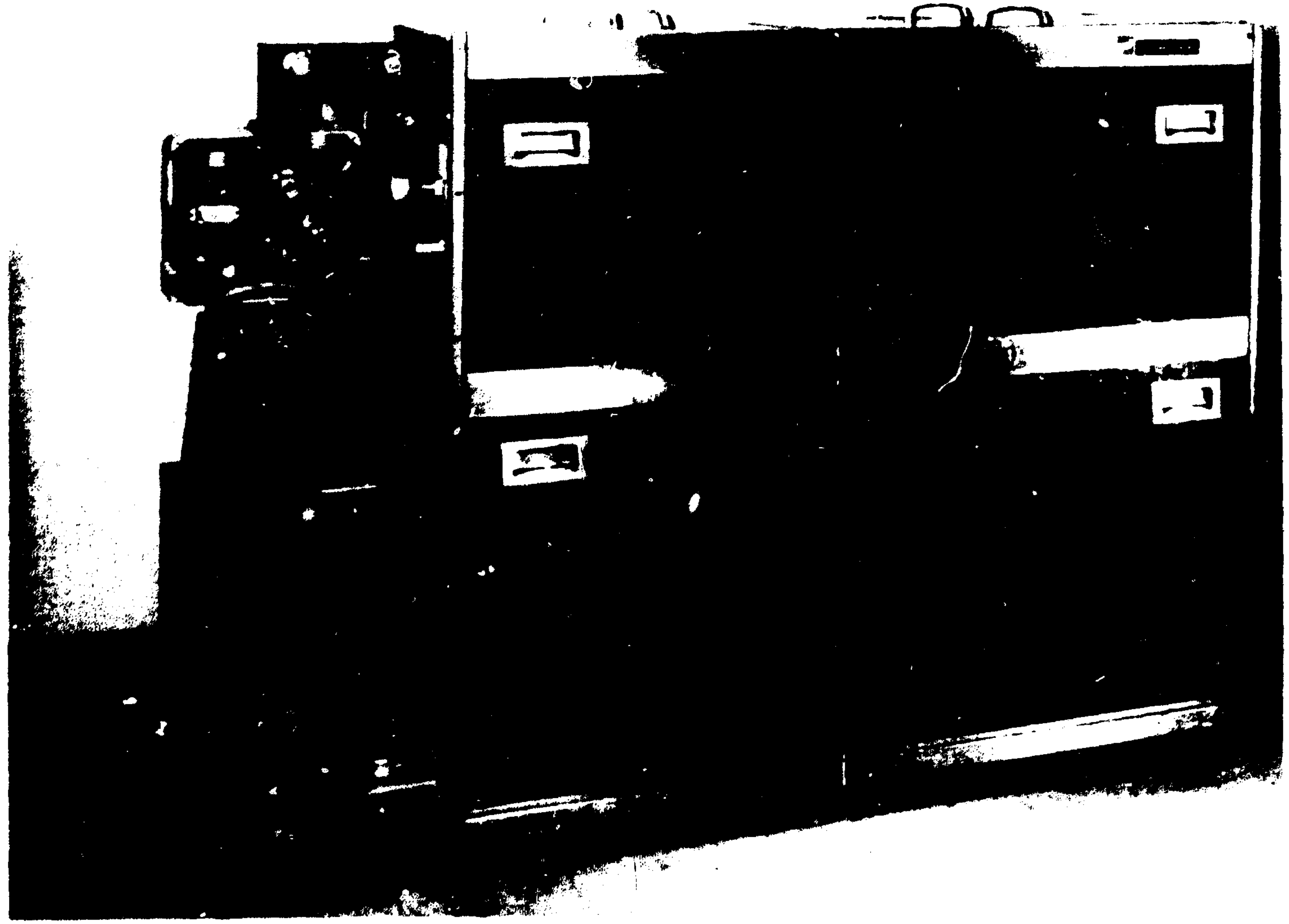


Fig. C-4 The Camera Processor

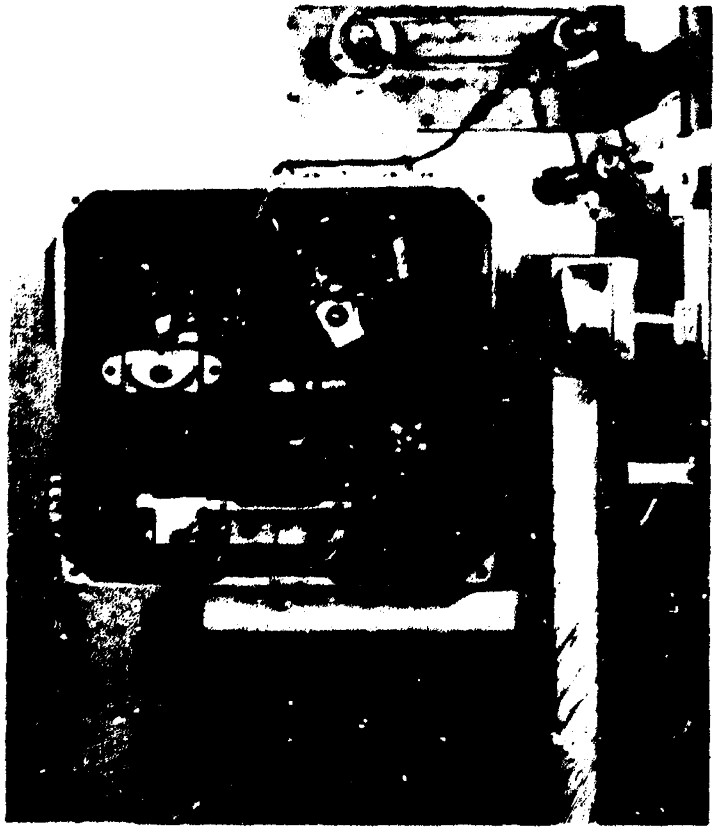


Fig. C-5 Camera and Transport Mechanism-Driving Side

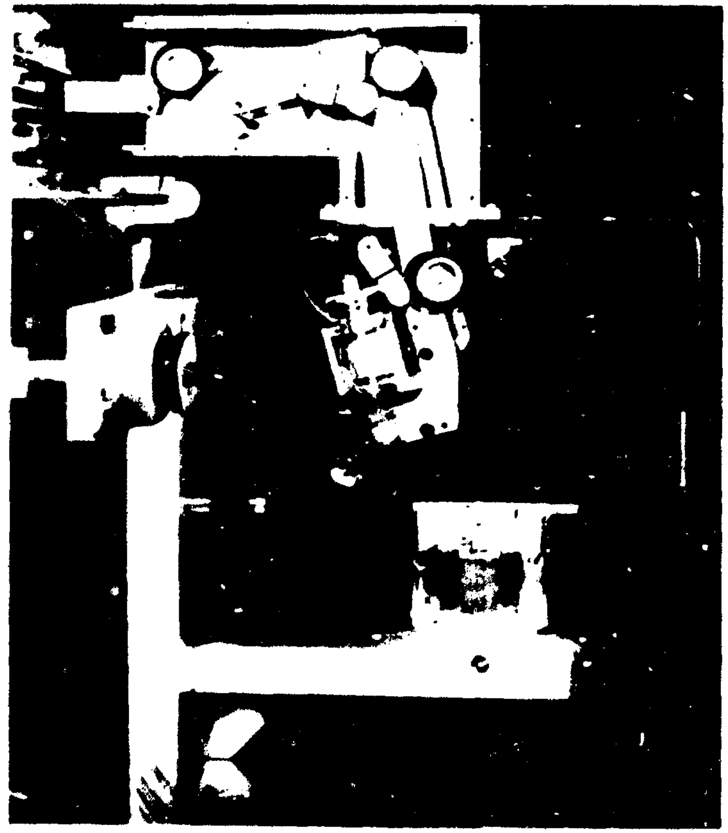


Fig. C-6 Camera and Transport Mechanism-Film Loading Side

These sense switches are connected with the film-terminal control-logic in such a manner as to prevent successive film strips from overlapping.

Figure C-6 is a photograph of the take-up chamber of the Kodak film unit; it shows the modifications which have been made to cut off automatically short lengths of film and transport them to the processor. At the end of a series of exposures, or after 10 adjacent frames have been exposed, the camera is commanded to advance the film until the last exposed frame has cleared the cutter. The solenoid-operated film-cutter then severs the exposed film which is then moved through the film guides to the processor by means of three neoprene drive-rollers.

Additional modifications were made to the Kodak film unit to hold the receiver lens in a precise position with respect to the CRT and film. The camera and CRT were mounted on a one-inch thick aluminum plate which was in turn bolted to the side of the GAF processor using shock-absorbing mountings.

The operation of the camera-processor is initiated by a digital signal transmitted from the central station. The actual sequencing of the camera-processor is accomplished by the film-terminal control circuitry. The logic for this circuitry is constructed from DTL and TTL dual-in-line integrated circuits. The camera solenoid, the film-cutter solenoid, and the control relay in the processor require approximately five amps and this current is switched with silicon controlled rectifiers.

To insure that textual images will be transmitted to the film terminal only when the camera shutter is open, the camera is loaded with film, and the CRT is operative, a status level will be transmitted to the central station.

In order to conserve water and mechanical wear on the GAF processor, a unijunction-transistor timer has been constructed to cycle the processor in the following manner. Immediately after the initial turn-on of the receiver station, the processor will run for 17 minutes to warm up. Following this initial warm-up period, the processor will be switched to a stand-by condition until a request to process a film strip has been received. If 15 minutes elapses before the next request, the processor will be run for two minutes to maintain the solution temperatures, and then return to the stand-by condition.

As reported in the preceding Semiannual Activity Report, the success of the microfilm-facsimile terminal depends upon the availability of convenient microfilm viewers for the film strips that emerge from the camera-processor unit. A prototype of a potentially-inexpensive, plastic film-strip holder has been constructed and a device that will accept this holder is being installed in an IBM-9922 document viewer.

7. TEXT-ACCESS SYSTEM CONTROL

The text-access system control (TASC) electronics provides the interface with the buffer/controller, the CARD unit, and the transmission subsystem. It receives messages of one, two, or three 16-bit words from the buffer/controller in the format shown in Fig. C-7. Upon receiving a user request through the

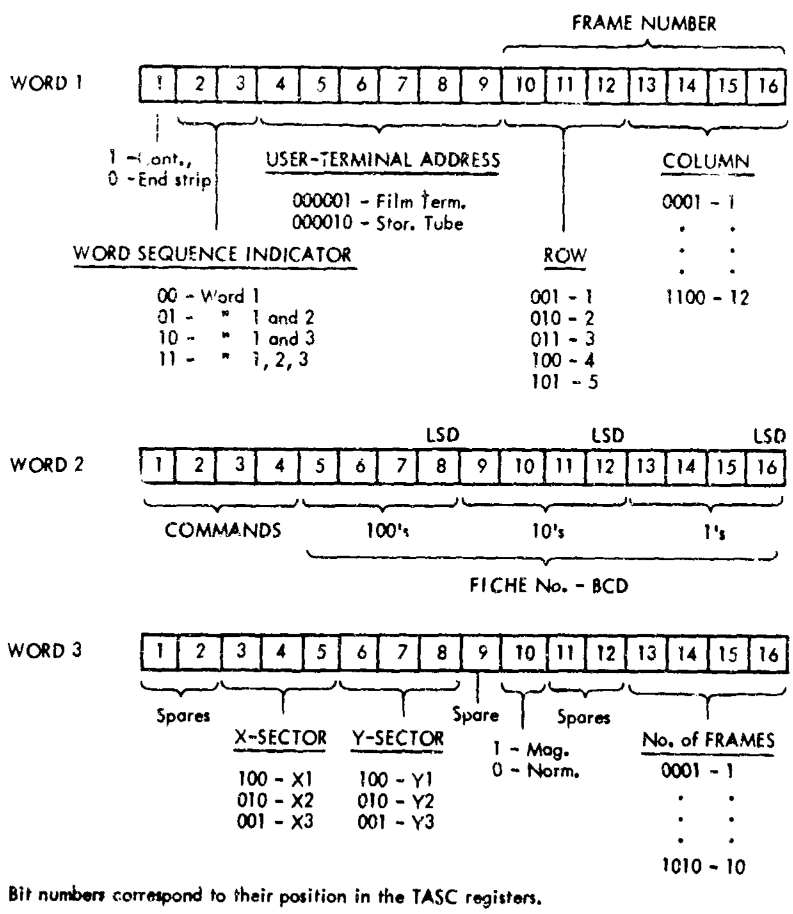


Fig. C-7 Input-Message Format

augmented-catalog console, the buffer/controller formats the message and sends only the words necessary to process the request. The first word is always required because it contains the user-terminal address to which the requested text will be transmitted. The second word is sent whenever the request requires a retrieval cycle in the CARD unit or when digital commands are transmitted to a user terminal. The third word is sent when the request is for a multiple number of frames at the film terminal or for a magnified display at the transient-display terminal.

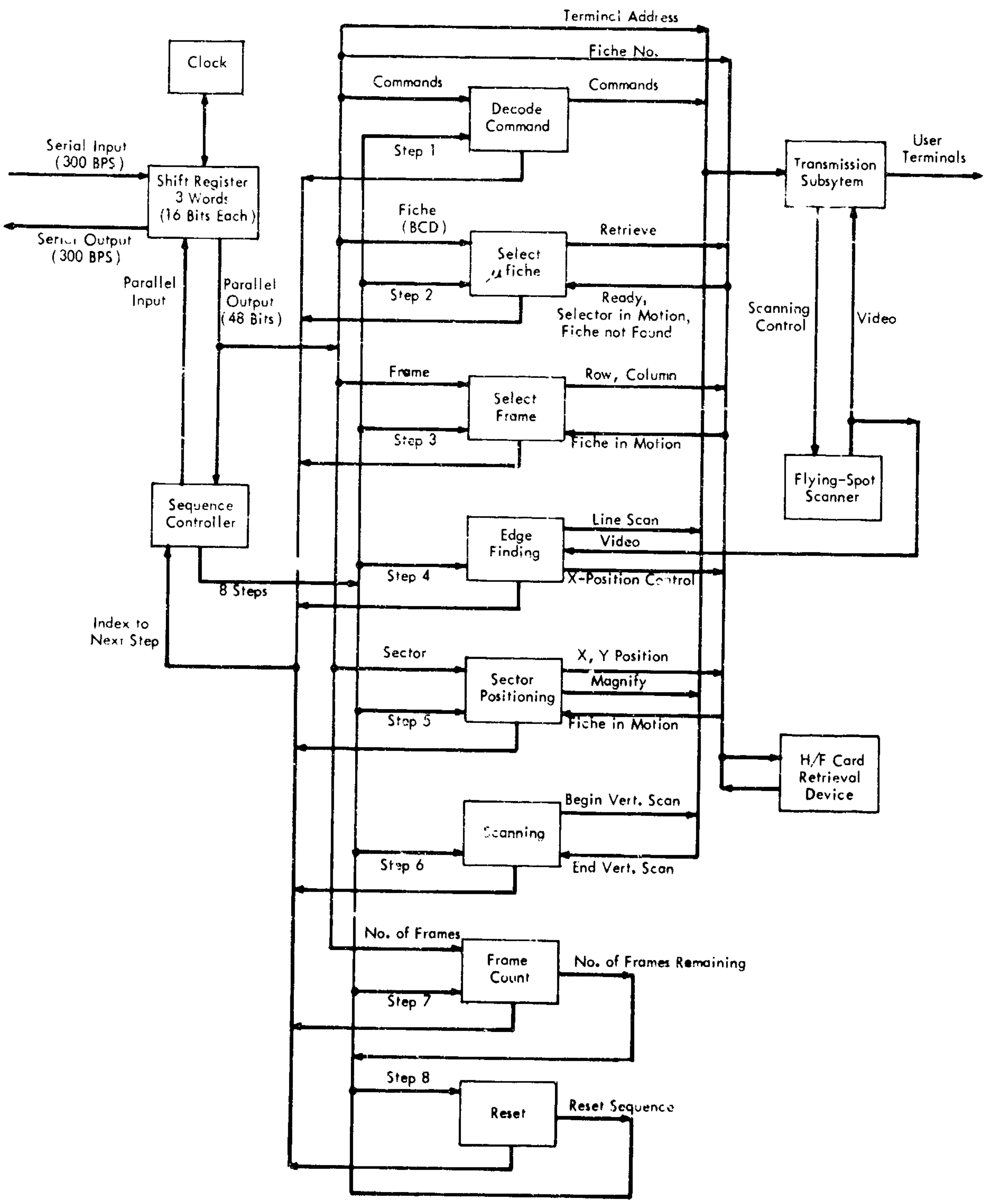


Fig. C-8 TASC Functional Diagram

The message format includes spare bits for additional functions that might be added to the system later. The command bits in the second word are intended for transmitting digitally coded information to the central station and the user terminals.

Communication between the buffer/controller and the TASC circuitry will be over a 300 bit-per-second serial data channel. Each 16-bit word will be preceded by an initial logical ZERO and followed by a logical ONE. The TASC electronics will contain three 16-bit shift and storage registers corresponding to word one, word two, and word three, and data from the buffer controller will be entered into the proper TASC shift register. After the completion of a request, or if a minor malfunction occurs, the registers will be cleared and a status message will be sent back to the buffer/controller.

A block diagram of the TASC electronics appears in Fig. C-8. The sequence controller is a modulo-9 counter which is indexed by signals indicting the completion of each step in the microfiche retrieval, frame positioning, and scanning cycles. This concept was chosen because it allows independent design of the logic for each step; independent design, in turn, simplifies not only the initial circuitry but also future modifications to the TASC electronics. Industrial-grade, dual-in-line integrated circuits and medium-scale-integration devices have been selected to implement the logic design. These components are mounted on double-sided, copper-clad, 4 x 6 inch circuit boards, as shown in Fig. C-9. The circuit boards are mounted in a standard 19-inch rack which is assembled into a cabinet, (Fig. C-10).

When the shift registers have received a complete message, a signal is generated which sets the sequence controller to step 1. If word two was included in the message, the step-1 logic will decode the command bits. Legal commands consist of the binary coded numbers 0 through 7. It is planned that these commands will be utilized to switch into a character-writing mode where digital characters may be sent to the film terminal to identify the film strips with notations readable by the unaided eye. If there is no second word or if the command 0 is sent then the sequence controller will skip to step 2.

The microfiche-retrieval cycle is accomplished during step 2. At the end of each scanning cycle, the microfiche remains in the CARD positioning device until the next request; a retrieval cycle is necessary only if the request involves a different fiche. If the request is for a frame on the same fiche as the previous transmission, the second word is not included in the message and the sequence is indexed through this step. If a new fiche must be retrieved, the buffer/controller includes the second word in its message. The fiche number of three digits in binary-coded decimal form is forwarded from the shift register directly to the CARD unit. At step 2 of the sequence, a retrieve signal to the CARD unit initiates the fiche-selection cycle. Feedback signals from the CARD unit indicate the end of the cycle and index the sequencer to the next step.

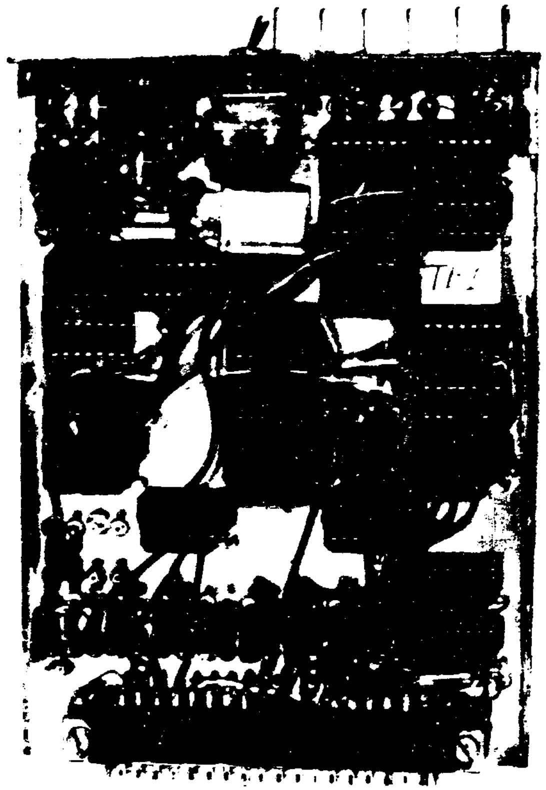


Fig. C-9 TASC Circuit Board

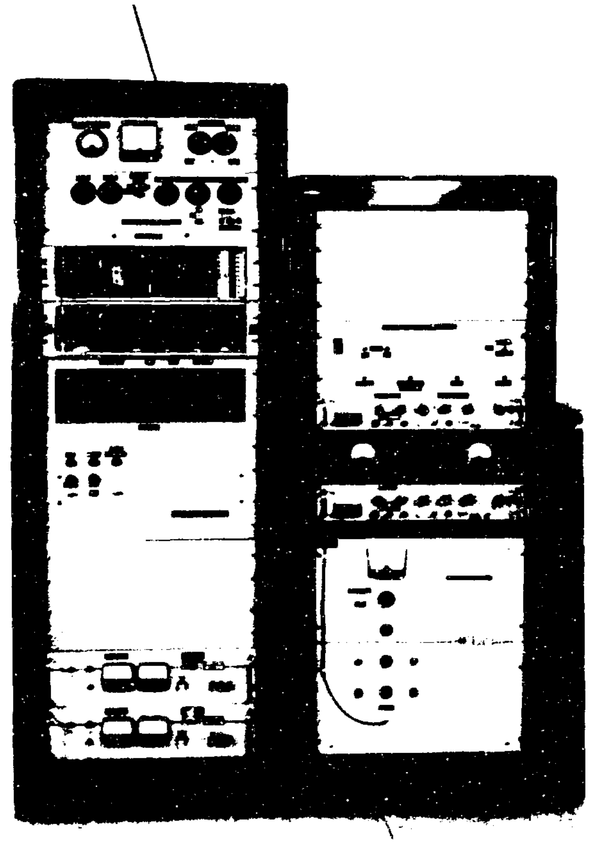


Fig. C-10 Text-Access Equipment Cabinet

In step 3, the x-y positioning servo of the CARD unit positions the requested frame in front of the scanner aperture. The binary-coded row and column information in the input message is decoded and activates one-of-five row inputs and one-of-twelve column inputs to the CARD unit. When the position servos reach the commanded positions, a pulse is fed back to index the sequence controller to the next step.

The tolerances for positioning the page images on the fiche during filming and fiche-assembling operations are such that the page locations can vary considerably with respect to the COSATI grid, particularly in the horizontal direction of the text. To center each page on the optical axis of the scanner, a technique for locating the left edge of the page image before scanning will be included in the positioning cycle. The technique is to use the frame positioning of step 3 for approximate positioning, and then in step 4 to scan the flying-spot scanner along a vertical line just outside the normal raster. The fiche is moved in a horizontal direction by the CARD servo until the edge of the page image is detected from the video signal. All filmed page images are separated by a clear gap and each has a dark border surrounding the text. The effect on the video signal of the transition from the gap to the border is used to detect the alignment of the page edge with the scanned line. The initial system will utilize the edge-finding technique only in the horizontal direction because it is believed that the vertical positioning of the pages on the fiche is sufficiently accurate to make a centering cycle in that direction unnecessary. When the edge is detected, the positioning servos are disabled and the sequence is indexed to the next step. The edge finder will be implemented during the next report period.

Step 5 is used only when a magnified display is requested at the transient-display terminal. As previously described, the user can select a 2X magnification of a quarter-page centered on one of nine points. During this step, the x- and y-sector bits in word 3 cause signals to be sent to the CARD servos which offset the fiche with respect to the scanner axis. During transmission, the scanner raster is reduced to one-half its horizontal and vertical dimensions which causes the characters in the reconstructed image at the display terminal to be twice their normal size. By shifting the fiche instead of the raster, the scanner is used as close to its optical axis as possible, thereby maximizing its resolution. A pulse at the end of the fiche-in-motion signal is used to indicate that the fiche is positioned at the requested sector and the sequence controller is indexed to the next step.

If the transient-display terminal is addressed, a command to begin vertical sweep is given to the transmitter logic when the sequence controller activates step 6. When the film terminal is addressed, the TASC logic monitors the film-terminal status line and will send images to the film terminal only if the status line is high. When the end of the vertical scan is received, the sequence controller will index to step 7.

If a user requests several pages from a document to be transmitted to the film terminal, the buffer/controller includes the number of pages in word 3 of the message,

and this number is loaded into a frame counter during step 1. At step 6, the counter is indexed down by one when a frame is transmitted. At step 7, the frame counter is checked to see if there are any remaining pages of the document to be transmitted, and if so, the sequence controller resets the sequence to step 3. The registers containing the frame row and column are indexed to move to the next page and the sequence continues from step 3. When the frame counter reaches zero at step 7, the last page of the request has been transmitted and a CUT-AND-PROCESS-FILM command is sent to the terminal and the sequence counter is indexed to step 8.

Upon entering step 8 the sequence counter is immediately reset to step 0, the resting state, and a REQUEST-COMPLETED message is transmitted to the computer.

In order to prevent the sequence controller from stopping due to an illegal message or minor equipment malfunction, a unijunction timer is started at the initiation of each step. If more than 15 seconds elapses without indexing from one step to another, the unijunction timer will trip,reset the sequence counter,and transmit an error message to the computer.

8. TEXT-ACCESS SOFTWARE

The text-access hardware system, described in the preceding sections, receives its input signals from the buffer/controller of the INTREX-catalog console. These input signals are generated within the buffer/controller by the 620I computer in response to user requests. As a consequence, certain software requirements are imposed on the 620I in order that it may select and format the command signals properly. These software features are described below.

Figure C-11 shows the buffer/controller in relation to the rest of the INTREX system. The 620I serves as a functional interface between the user and the text-access controller. A set of function switches located at either the display station or at the catalog console may be used as input devices to the text-access system. The 620I formats the proper command sequences and issues them to the text-access controller. The system is designed for multiple users and the 620I maintains the queueing algorithms for text-access requests.

Much attention has been given to the problem of communication between the 620I computer and the text-access controller. Flexible operation is desirable with a minimum of information transmission between devices. The following discussion of the control-and-communication features is given from the viewpoint of the 620I programmer who regards the text-access controller as another I/O device.

The 620I communicates with the text-access controller through a 16-bit word buffer attached to the I/O bus of the computer, as shown in Fig. C-12. Under program control 16-bit computer words can be transferred to, or from, the text-access controller and the internal registers and core storage of the 620I computer. Words sent to

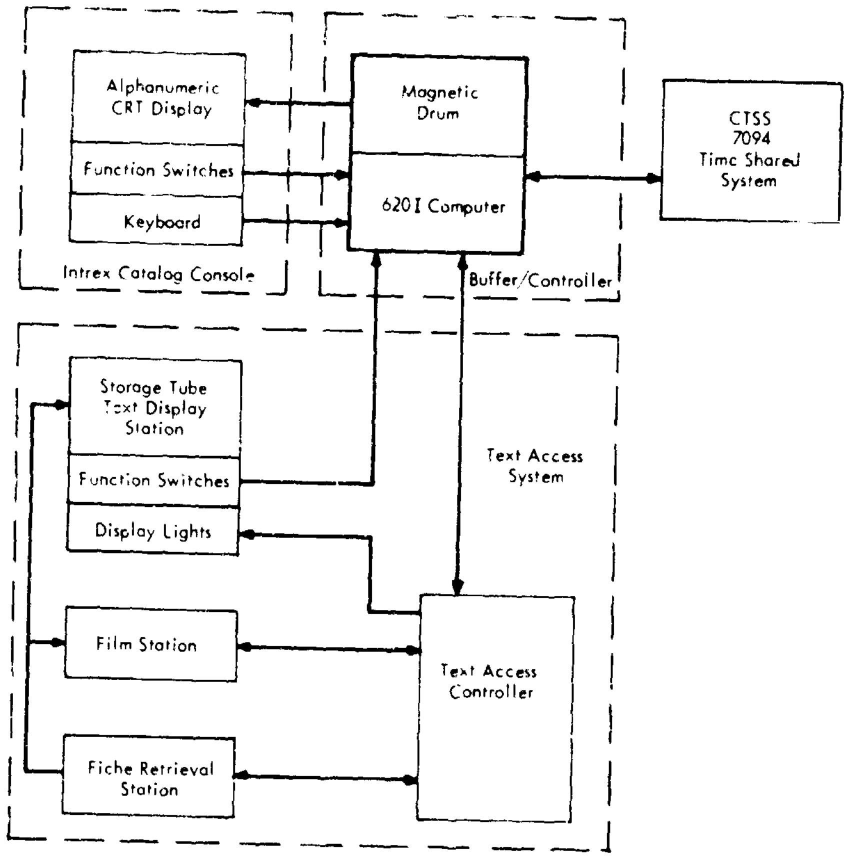


Fig. C-11 Intrex System Configuration

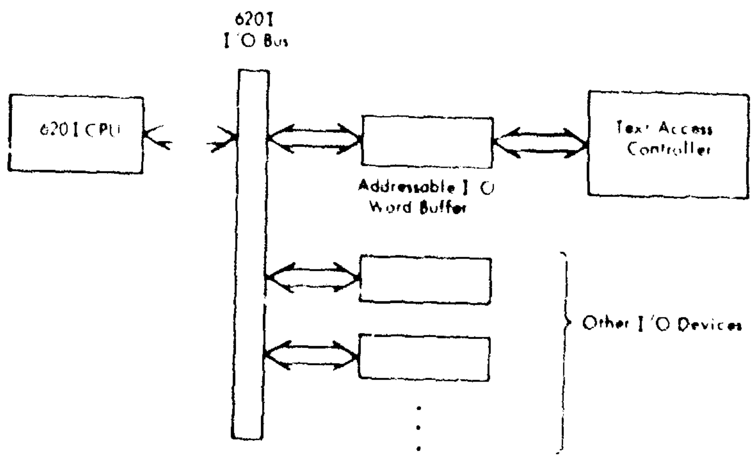


Fig. C-12 Interface of Text-Access Controller with 6201 Computer

the text-access controller serve as commands to the film and display stations. The word elements of the text-access commands are:

Terminal Address: A 6-bit device address indicating the device to which the command sequence is directed. The film station is assigned Device Address 1 and the display station, Device Address 2. There is addressing capability for 63 devices.

Fiche and Frame Numbers: The retrieval device requires a fiche number in order to select a document from the store and a frame number in order to position a page for either viewing or film reproduction.

Fiche Number: A 12-bit number consisting of 3 BCD digits, each containing four bits. The fiche numbers may range from 000 to 999 although zero does not correspond to a stored fiche and the retrieval device may not have a capacity of 999. The fiche number corresponds to the unit record of the retrieval device. That record consists of 60 separate frames. Each frame normally corresponds to a single page of the filmed document.

Frame Number: A 7-bit number consisting of a 3-bit-row number and a 4-bit-column number. This arrangement is shown schematically in Fig. C-13.

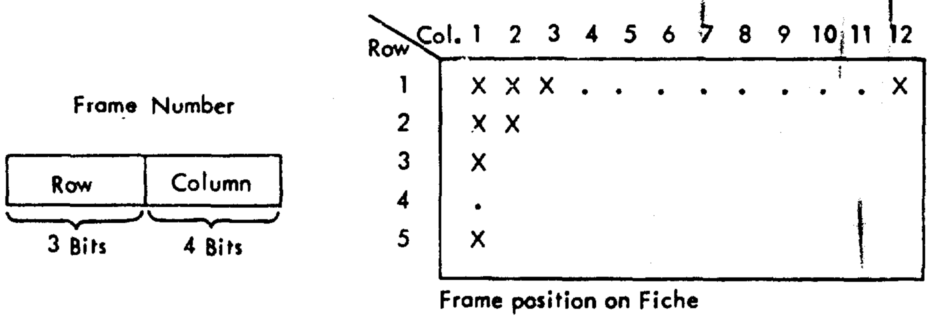


Fig. C-13 Frame Number and Corresponding Fiche Position

The fiche contains 60 frames arranged in 5 rows and 12 columns. The (1, 1) position would be referenced by the binary string 001 0001; the (5, 12) position by 101 1100. This code was imposed by the retrieval-device design.

Number of Frames: A 4-bit number which indicates the number of frames to be placed on one strip of the reproduced film generated by the film station when a copy is requested. Currently hardware requirements set a maximum of 10 frames per strip. If the document exceeds 10 pages, 620I Program-Control issues the number of film requests needed to copy the entire document. Program-Control will also set a limit on the number of pages a user may request at a given time.

Strip Continue: A 1-bit indicator for the film station. If the bit is set to one, the film station does not cut the film after completing the current copy request. The user is thus allowed to place more than one document on the same film strip. The usual situation will be to cut the film strip after a given document has been reproduced, and so the Strip Continue bit would be zero.

Magnify: A 1-bit indicator requesting that a displayed image be magnified.

X and Y Sector Numbers: Two 3-bit numbers represented in a one-out-of-three code used to give the X, Y coordinates of the portion of the page (called the sector) to be magnified. Figure C-14 presents the coordinate representation. The one-out-of-three code is selected for convenience of decoding at the text-access controller.

Y \ X	001	010	100
001	S1	S2	S3
010	S4	S5	S6
100	S7	S8	S9

Fig. C-14 X,Y Coordinate Representation of Page Sector for Magnification Option

Command: A 4-bit command code used for issuing commands either to the text-access controller or to the 620I buffer/controller. These special codes are included to allow flexible operation and system modification as the Intrex experiments proceed. They will also be used to convey status information from the text-access controller to the 620I.

Word-Sequence Indicator: A 2-bit code used to indicate to the text-access controller which of four different command-word sequences is currently being issued by the 620I. A detailed discussion of this feature follows in the next paragraphs.

COMMAND-WORD SEQUENCE

The action of the text-access controller is directed by the 620I which formats the above data elements into command words and outputs those command words via the I/O bus in a sequence determined by the system state at the time the request is processed. There are three command-word formats.

The first word contains the Text-Access-Device Address, the Frame Number, the Strip-Continue bit, and the Word-Sequence Indicator. Designated by the word

formats WORD1, WORD2 and WORD3, the Word-Sequence Indicator has the following code:

00 Only WORD1 will be sent for this command. The word sequence indicator itself is contained in WORD1.

01 WORD1 and WORD2 will be issued in that sequence for the current command.

10 WORD1 and WORD3 will be issued in that sequence for the current command.

11 WORD1, WORD2 and WORD3 will be issued in that sequence for the current command.

All commands require a WORD1.

The second word contains the 3-digit BCD fiche number and the 4-bit command code. The third word contains either the 4-bit number-of-pages count or the X and Y sector numbers with the magnify bit set to one. Figure C-7 summarizes the command-word-format information.

DISPLAY-COMMAND SEQUENCE

To command the text-access system to retrieve a specific page of a stored document and to display it at the storage-tube station, the 620I normally formats and outputs a 2-word sequence to the text-access controller. The first word contains the terminal address of the display station, the frame number (row and column) of the page to be displayed and a word-sequence indicator of 01. The strip continue indicator is not significant in this example and is ignored by the controller. The second word contains the 3-digit fiche number and the command code is zero. (A zero code is used for all display and copy commands. The terminal address determines which command is display and which is copy.) The text-access controller causes the retrieval device to select the requested fiche and to position the proper frame (page) for display on the storage-tube-display device.

The retrieval device does not re-store the selected fiche until a new request is made. If the next request is for another frame on the same fiche (as is often the case for a user viewing successive pages of a single document), then the 620I sends only WORD1 for each subsequent request for the same document. Currently, a single document is completely contained on one fiche. Such subsequent requests require a new frame number and the word-sequence indicator set to 00 to inform the text-access controller that only one command word is being sent for that request.

MAGNIFIED-DISPLAY OPTION

If the user has selected a page and has it displayed on the storage tube he may decide to have the disp' station present an enlarged display of part of that page. The 620I then formats and issues a WORD1, WORD3 sequence to the text-access controller. The only change in WORD1 from the previous request is that the word-sequence

indicator sets to 10 in order to flag a WORD1, WORD3 sequence. The terminal-and-frame information is unchanged; however, they must be re-issued.

WORD3 will contain the X- and Y-sector numbers and the magnify bit set to one. The number-of-frames part of the word is ignored by the controller.

ALTERATION OF COMMAND SEQUENCE - MULTIPLE USER

The text-access system is designed for multiple-user operation. Often, a second user may request a display or film copy of a document before user number 1 has finished paging through his currently-selected document or magnifying his currently displayed page. Since the retrieval device itself can only have one fiche selected at any given time, the first fiche must be restored before the second fiche can be retrieved. The first user still has his image to view and he can still request the next page or the magnification of his current page; however, this request cannot be honored until the second user's fiche has been retrieved, processed (filmed or displayed), and restored. Then the first fiche can be re-retrieved by the 620I through issuance of a 2-word sequence. since User 1's WORD2 was lost in the processing of User 2's request. For User 1 to employ the magnify option, the complete 3-word sequence is needed because of the intervening User-2 request. The 620I program control determines which word sequence is required under the existing circumstances.

FILM-REPRODUCTION-COMMAND SEQUENCE

To request a film copy of a document without having it displayed, a 3-word sequence is issued by the 620I computer. The only significant information in WORD3 is the number of frames (pages) to be filmed. Normally, this is the same as the number of pages in the document if its length is ten pages or less. WORD1 contains the terminal address of the film station which informs the text-access controller that the request is for filming and not for display. The Strip-Continue bit will normally be zero unless the user wants to place more than one document on the same film strip.

The more probable case will be that the user has viewed a document and then wishes to have it copied by the film station. If that fiche has remained selected because no other user has made an intervening request, then the 620I may issue only a WORD1, WORD3 sequence for filming. No WORD2 is needed. As the text-access controller is currently designed, the computer program must monitor the retrieval-station status and skip issuing WORD2 if the proper fiche is already selected. If that function is not performed by the 620I, then the fiche will be restored and reselected, thereby requiring a complete access cycle of the retrieval device.

MULTIPLE-USER PRIORITY

Multiple-user operation may cause queues to develop for both the display and filming functions. Filming a multiple-page document engages the retrieval device

for more time than the processing of a display request. From a user-convenience viewpoint, waiting for a page display would seem more frustrating than delaying the availability of a copied document. Hopefully, the copy will be processed by the time the user reaches the film-station location. Given these considerations a higher priority has been assigned to display requests than to film requests, and, therefore, the 620I scheduling algorithm will favor the display queue over the copy queue. Experiments with the system may indicate other requirements on the scheduling algorithm.

III. PROJECT INTREX STAFF

A. PROJECT OFFICE

Professor Carl F. J. Overhage, Director
Mr. Joseph J. Beard
Mr. Charles H. Stevens

B. ELECTRONIC SYSTEMS LABORATORY

Professor J. Francis Reintjes, Director
Mr. John E. Ward, Deputy Director
Mr. Alan R. Benenfeld
Mrs. Sonja K. Escudier
Dr. Uri F. Gronemann
Mrs. Elizabeth J. Gurley
Dr. Donald R. Haring
Mr. Charles E. Hurlburt
Mr. James E. Kehr
Mr. Donald R. Knudson
Mr. Peter Kugel
Mr. Robert L. Kusik
Mrs. Sondra F. Lage
Miss Lucy T. Lee
Mr. Richard S. Marcus
Mrs. Jane R. Marshall
Mr. Michael K. Molnar
Miss Sonia P. Niessen
Professor James K. Roberge
Mr. Peter R. Scott
Professor Alfred K. Susskind
Mr. Stephen N. Teicher
Mr. Herman F. Vandevenne

C. ENGINEERING LIBRARY

Miss Rebecca L. Taggart, Head
Miss Margaret W. Artinian
Miss Barbara C. Darling
Mrs. Jean Field
Mr. Jeffrey L. Gardner
Mr. James M. Kyed
Miss Helen Magedson
Mrs. Suanne W. Muehlner
Miss Susan Nutter
Miss Mary Pensyl
Mrs. Alice M. Robrish
Mrs. Colleen M. Scholz
Mrs. Ines Siscoe

IV. CURRENT PUBLICATIONS

A. REPORTS

Benenfeld, A. R., "Generation and Encoding of the Project Intrex Augmented Catalog Data Base." ESL-R-360, August, 1968. This report is based upon a paper presented at the 6th Annual Clinic on Library Applications of Data Processing, University of Illinois, Urbana, Illinois, May 7, 1968; it will also appear in the published proceedings of the conference.

Kampe, W. R., "Pre-Indexing by Machine," ESL-R-355, July, 1968.

Lovins, J. B., Miss, "Development of a Stemming Algorithm." ESL-TM-353, June, 1968.

B. JOURNAL ARTICLES AND CONFERENCE PAPERS

Haring, D. R., "A Display Console for an Experimental Computer-Based Augmented Library Catalog." 1968 ACM National Conference and Exposition. Las Vegas, Nevada, August 27-29, 1968. (To appear in Conference Proceedings.)

Reintjes, J. F., "Project Intrex, (A Research Project under Development)." Presented at the Conference on Computers and the University, Technical University of Berlin, July 22-August 2, 1968.

C. THESES

Kampe, W. R., "Pre-Indexing by Machine." M.S. Thesis, Electrical Engineering Department, Massachusetts Institute of Technology, June, 1968. Also Electronic Systems Laboratory Report ESL-R-355, Published July, 1968.

Lufkin, R. C., "Determination and Analysis of Some Parameters Affecting the Subject Indexing Process." B.S. Thesis, Massachusetts Institute of Technology, June, 1968.

D. MISCELLANEOUS PRESENTATIONS

Carl F. J. Overhage

"Plans for Project Intrex." Symposium on Mechanized Information Services in the Library, UCLA, March 29-31, 1968.

"Information Transfer in the Library of the Future." Panel of the International Communications Conference on "The Evolving Relationship Between the Publisher and the Electronic Firm --Software and Hardware," IEEE, Philadelphia, June 12, 1968.

"Information Transfer Experiments at M.I.T." International Federation for Information Processing IFIP 68 Congress, Edinburgh, Scotland, August 6, 1968.

Charles H. Stevens

"Engineering Libraries and Information Systems." Center for Advanced Engineering Study, M.I.T., April 16, 1968.

"Books and Other Paraphernalia - A Forward Look." National Endowment for the Humanities, Boston, May 10, 1968.

"Console Development at Project Intrex." Special Libraries Association, Los Angeles, June 5, 1968.

V. PAST PUBLICATIONS

A. REPORTS

Benenfeld, A. R., Gurley, E. J., and Rust, J. E., "Cataloging Manual." ESL-TM-303, February, 1967.

Gronemann, U. F., Knudson, D. R., and Teicher, S. N., "Remote Text Access for Project Intrex." ESL-TM-312, July, 1967. This report is based upon a conference paper presented at the National Microfilm Association Convention, Miami Beach, Florida, April 26-28, 1967.

Haring, D. R., and Roberge, J. K., "The Augmented Catalog Console for Project Intrex." ESL-TM-323, October, 1967.

Project Intrex Staff, Semiannual Activity Report, 15 March 1966.

Project Intrex Staff, Semiannual Activity Report, 15 September 1966.

Project Intrex Staff, Semiannual Activity Report, 15 March 1967.

Project Intrex Staff, Semiannual Activity Report, 15 September 1967.

Project Intrex Staff, Semiannual Activity Report, 15 March 1968.

B. JOURNAL ARTICLES AND CONFERENCE PAPERS

Benenfeld, A. R., "Data Encoding for the Project Intrex Augmented-Catalog Experiments." Presented at the American Documentation Institute Convention, User Discussion Group on Emerging Machine-Readable Tape Services, New York, October 25, 1967.

Haring, D. R., "A Terminal for an On-Line Interactive Retrieval System." Presented at the IEEE Workshop on Advanced Computer Peripherals, Lake Arrowhead, California, August 25-27, 1967.

Overhage, Carl F. J., "Plans for Project Intrex." Science, 152, 20 May 1966, pp. 1032-1037.

Overhage, Carl F. J., "Science Libraries: Prospects and Problems." Science, 155, 17 February 1967, pp. 802-806.

C. THESES

Domercq, R. J., "A Machine-Aided Thesaurus Generation System." M.S. Thesis, Massachusetts Institute of Technology, September, 1967.

Stuntz, S. C., "Inputting Process to an Augmented Library Catalog -- An Efficiency Study." B.S. Thesis, Massachusetts Institute of Technology, June, 1967.

D. MISCELLANEOUS PRESENTATIONS

Gary L. Benton

"Project Intrex - Information Transfer Experiments at M.I.T." Raytheon Institute of Bedford, Bedford, Mass., February 15, 1966

Panel on New Hardware Developments, ADI Annual Meeting, Santa Monica, California, October 5, 1966.

"Project Intrex: Plans and Progress." Massachusetts Association of Medical Record Librarians, Boston, Mass., January 26, 1966.

Carl F. J. Overhage

"Plans for Information Transfer Experiments at M.I.T."

American Documentation Institute, Cambridge, Mass., October 28, 1965.

American Association for the Advancement of Science, Berkeley, California, December 29, 1965.

American Library Association, Chicago, Illinois, January 26, 1966.

American Physical Society, New York, N.Y., January 2 , 1966.

M.I.T. Lincoln Laboratory, Lexington, Mass., February 8, 1966.

Optical Society of America and American Chemical Society, (Rochester sections) Rochester, N.Y., March 1, 1966.

Rochester Computer Systems Association, Rochester, N.Y., April 26, 1966.

AFIPS Spring Joint Computer Conference, Boston, Mass., April 28, 1966.

Association of Research Libraries, New York, N.Y., July 9, 1966.

Panel on Information Systems Applications, ADI Annual Meeting, Santa Monica, California, October 5, 1966.

Panel on National Information Issues and Trends, ADI Annual Meeting, Santa Monica, California, October 6, 1966.

"Plans for Project Intrex at M.I.T."

M.I.T. Sloan School of Management, Cambridge, Mass., October 13, 1966.

American Cyanamid Company, Librarians and Information Specialists Meeting, Princeton, N.J., October 20, 1966.

M.I.T. Center for Advanced Engineering Study, Cambridge, Mass., November 29, 1966.

"Project Intrex - Plans and Progress." Harvard University Librarian's Seminar, Cambridge, Mass., May 17, 1967.

"Project Intrex and the Future of the Book." American Book Publishers' Association, Arden House, Harriman, N.Y., March 16, 1967.

Charles H. Stevens

"Project Intrex." Special Libraries Association (Boston Chapter), Boston, Mass., January 17, 1966.

"The Future of the Book." Boston Bookbuilders Workshop, Boston, Mass., March 16, 1966.

"Project Intrex at M.I.T." Simmons College Library School Alumni Association, Boston, Mass., April 12, 1966.

"Future Trends in Academic Libraries." Simmons College Library School, Boston, Mass., April 22, 1966.

"Project Intrex and the Special Library." Special Libraries Association, Minneapolis, Minn., May 30, 1966.

"Special Libraries in the Future." John Cotton Dana Lecture, University of Texas, Austin, Texas, September 23, 1966.

"Project Intrex: Plans and Progress." Special Libraries Association - Texas Chapter, Austin, Texas, September 24, 1966.

"Project Intrex and Information Transfer." Sloan Fellows Seminar, M.I.T., Cambridge, Mass., October 13, 1966.

"Project Intrex and the Future of the Academic Library." Science and Technology Information Group, Washington, D.C., October 25, 1966.

"Project Intrex: Plans and Progress." Information Sciences Colloquium, Lehigh University, Bethlehem, Pa., October 27, 1966.

"Libraries in the Age of Electronics." Library Symposium, Boston University, Boston, Mass., November [illegible]9, 1966.

"New Developments in Libraries." State University of New York (Albany), April 4, 1967.

"Project Intrex." Gordon Research Conference on Scientific Information, New London, New Hampshire, July 19, 1967.

"Project Intrex - Plans and Progress." New York Chapter, Special Libraries Association, New York, February 6, 1968.